

**THE ASSOCIATION OF SINGLE NUCLEOTIDE POLYMORPHISMS IN  
INTRONIC REGIONS OF ISLET CELL AUTOANTIGEN 1 AND TYPE 1 DIABETES  
MELLITUS**

by

**Carrie Lynn Blout**

B.S., Dickinson College, 2004

Submitted to the Graduate Faculty of  
Department of Human Genetics – Genetic Counseling  
the University of Pittsburgh in partial fulfillment  
of the requirements for the degree of  
Masters of Science

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH  
THE GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Carrie Lynn Blout

It was defended on

March 30, 2006

and approved by

**Committee Member:**

Elizabeth Gettig, MS, CGC  
Associate Professor of Human Genetics  
Director, Genetic Counseling Program  
Graduate School of Public Health  
University of Pittsburgh

**Committee Member:**

John W. Wilson, PhD  
Assistant Professor  
Department of Biostatistics  
Graduate School of Public Health  
University of Pittsburgh

**Committee Member:**

Michael M. Barmada, Ph.D.  
Associate Professor of Human Genetics  
Director, Human Genetics Department Computational Resources Division  
Graduate School of Public Health  
University of Pittsburgh

**Thesis Advisor:**

Massimo Pietropaolo, M.D.  
University of Pittsburgh Physicians Faculty UPMC  
Department of Pediatrics  
Rangos Research Center Genetic Immunology  
University of Pittsburgh

# **THE ASSOCIATION OF SINGLE NUCLEOTIDE POLYMORPHISMS, LOCATED WITHIN INTRONIC REGIONS OF ISLET CELL AUTOANTIGEN 1 AND TYPE 1 DIABETES MELLITUS**

Carrie L. Blout, M.S.

University of Pittsburgh, 2006

Type 1 diabetes mellitus is a multifactorial autoimmune disease caused by a combination of genetic and environmental factors. Further knowledge and understanding about the genes which play a role in type 1 diabetes has a clear public health significance in that it will aid in the prediction, treatment and a possible cure. Type 1 diabetes is a chronic disease with a lengthy preclinical course, which eventually results in pancreatic beta cell destruction and inability of the body to produce insulin hormone. Type 1 diabetic patients generally require exogenous insulin to survive. Several genetic loci have been proposed to be linked to type 1 diabetes; however, the HLA and VNTR regions are currently believed to account for the majority of genetic risk, contributing to about 42% and 10% of an individual's risk to develop type 1 diabetes, respectively. This study focuses on the candidate gene *Islet Cell Autoantigen 1 (ICA1)*, which codes for the protein ICA69. This protein product is expressed in the islets of Langerhans, the neuroendocrine system and in the thymic medulla; this last location is an area of the body known to play a major role in immunologic tolerance. Preliminary studies in the NOD and B6 mouse models suggest that a SNP within the promoter region of *Ica1* affects transcription and may account for altered expression in the thymus. Our current study aimed to determine whether single nucleotide polymorphisms within intronic regions of the human *ICA1* gene differed between a diabetic case population and non-diabetic controls. It was hypothesized that SNPs within the *ICA1* gene differ between cases and controls and play a role in the onset of type 1 diabetes. Polymerase Chain Reaction (PCR) was applied for DNA amplification and the pyrosequencing technique was used to genotype all samples. At SNP location rs2058519 there was a clear genotypic difference between the cases and controls ( $p = .0003$ ). These results suggest that genetic variation at this specific SNP location in the *ICA1* may be associated with type 1 diabetes susceptibility. The ultimate goal of this study is to determine whether our candidate gene *ICA1* appears to play a role in the pathogenesis of type 1 diabetes.

## TABLE OF CONTENTS

|  |           |
|--|-----------|
| <b>ACKNOWLEDGEMENTS .....</b>                              | <b>IX</b> |
| <b>1.0 INTRODUCTION.....</b>                               | <b>1</b>  |
| <b>1.1 PURPOSE OF THE STUDY .....</b>                      | <b>1</b>  |
| <b>2.0 BACKGROUND AND SIGNIFICANCE .....</b>               | <b>2</b>  |
| <b>2.1 DIABETES OVERVIEW .....</b>                         | <b>2</b>  |
| <b>2.2 ENVIRONMENTAL FACTORS.....</b>                      | <b>3</b>  |
| <b>2.3 PHYSIOLOGY .....</b>                                | <b>4</b>  |
| <b>2.4 GENETIC SUSCEPTIBILITY .....</b>                    | <b>5</b>  |
| <b>2.5 ROLE OF THE HLA COMPLEX.....</b>                    | <b>8</b>  |
| <b>2.6 ROLE OF THE INSULIN GENE .....</b>                  | <b>11</b> |
| <b>2.7 IMMUNE SYSTEM .....</b>                             | <b>12</b> |
| <b>2.8 AUTOANTIBODIES.....</b>                             | <b>15</b> |
| <b>2.9 CANDIDATE GENE APPROACH .....</b>                   | <b>16</b> |
| <b>2.10 ISLET CELL AUTOANTIGEN 69KDA (ICA69).....</b>      | <b>18</b> |
| <b>2.11 THE PYROSEQUENCING METHOD .....</b>                | <b>20</b> |
| <b>2.12 SPECIFIC AIMS .....</b>                            | <b>22</b> |
| <b>3.0 MATERIALS AND METHODS .....</b>                     | <b>23</b> |
| <b>3.1 STUDY POPULATION .....</b>                          | <b>23</b> |
| <b>3.2 SNP IDENTIFICATION.....</b>                         | <b>24</b> |
| <b>3.3 PRIMER DESIGN AND TESTING .....</b>                 | <b>25</b> |
| <b>3.4 SNP LOCATIONS.....</b>                              | <b>27</b> |
| <b>3.5 PCR AMPLIFICATION FOR SEQUENCING PRODUCTS .....</b> | <b>28</b> |
| <b>3.6 DNA SEQUENCING AND GENOTYPE DETERMINATION.....</b>  | <b>30</b> |
| <b>3.7 STATISTICAL METHODS .....</b>                       | <b>34</b> |

|             |   |           |
|-------------|---|-----------|
| <b>4.0</b>  | <b>RESULTS .....</b>  | <b>36</b> |
| <b>4.1</b>  | <b>GENOTYPE COMPARISON AT THREE SNP LOCI.....</b>                             | <b>36</b> |
| <b>T4.2</b> | <b>HWE ANALYSIS OF GENOTYPING DATA .....</b>                                  | <b>41</b> |
| <b>5.0</b>  | <b>DISCUSSION .....</b>   | <b>42</b> |
| <b>5.1</b>  | <b>SNP SIGNIFICANCE .....</b>   | <b>42</b> |
| <b>5.2</b>  | <b>POSSIBLE FUNCTIONAL RELATIONSHIP OF AN INTRONIC<br/>POLYMORPHISM .....</b> | <b>42</b> |
| <b>5.3</b>  | <b>IDEAL SAMPLE SIZE.....</b>   | <b>43</b> |
| <b>5.4</b>  | <b>FUTURE STUDIES/APPROACH.....</b>   | <b>44</b> |
| <b>5.5</b>  | <b>GENOTYPING TROUBLE SHOOTING.....</b>                                       | <b>45</b> |
| <b>5.6</b>  | <b>OTHER POSSIBLE TECHNIQUES.....</b>   | <b>46</b> |
| <b>5.7</b>  | <b>SUMMARY .....</b>  | <b>46</b> |
|             | <b>BIBLIOGRAPHY .....</b>   | <b>48</b> |

## LIST OF TABLES

|  |    |
|--|----|
| Table 1. Empiric risks of type 1 diabetes. These estimates are for North American Caucasian* and Scandinavian populations**, ‡ 1-15% range depending on the populations (10-15)..... | 6  |
| Table 2. ICA1 SNPs considered to be ‘true’ which were previously validated or found in multiple databases.....   | 24 |
| Table 3. ICA1 SNPs which were not used in this study because they were not previously validated and were identified in only a single database.....                                   | 24 |
| Table 4. SNP locations within the ICA1 gene.....   | 25 |
| Table 5. PCR and Sequencing Primers for each SNP.....  | 30 |
| Table 6. Pyrosequencing Assay Information.....   | 32 |
| Table 7. Genotype data for rs3807830 in diabetics and controls.....  | 36 |
| Table 8. Genotype data for rs2348895 in diabetics and controls.....  | 36 |
| Table 9. Genotype data for rs2058519 in diabetics and controls.....  | 37 |
| Table 10. Comparison of original diabetic samples versus additional diabetic samples for SNP rs3807830.....  | 37 |
| Table 11. Comparison of original diabetic samples versus additional diabetic samples for SNP rs2348895.....  | 37 |
| Table 12. Comparison of original diabetic samples versus additional diabetic samples for SNP rs2058519.....  | 38 |
| Table 13. 3x2 Table comparison of pooled cases vs control samples for SNP rs3807830. ....  | 38 |
| Table 14. 3x2 Table comparison of pooled diabetic cases vs control samples for SNP rs2348895. ....   | 39 |
| Table 15. 3x2 Table comparison of pooled diabetic cases vs control samples for SNP rs2058519. ....   | 39 |

|  |    |
|--|----|
| Table 16. 2x2 Table comparing individuals with the common homozygote vs the rare homozygote or heterozygous individuals for SNP rs3807830. ....            | 40 |
| Table 17. 2x2 Table comparing individuals with the common homozygote vs the rare homozygote or heterozygous individuals for SNP rs2058519. ....            | 40 |
| Table 18. 2x2 Table comparing individuals with the common homozygote or heterozygous individuals vs those with the rare homozygote for SNP rs2058519. .... | 40 |
| Table 19. 2x2 Table comparing individuals with the common homozygote or heterozygous individuals vs those with the rare homozygote for SNP rs2058519. .... | 41 |
| Table 20. HWE chi-squared p-values for the three SNPs studied. ....  | 41 |

## LIST OF FIGURES

|   |    |
|---|----|
| Figure 1. Glucose Induced Insulin secretion from the pancreatic beta cells. Pietropaolo M, Roith Derek. (2001) Clinical cornerstone. Pathogenesis of Diabetes: Our Current Understanding. 4 (2): 1-6. ....  | 4  |
| Figure 2. General hypothesis for the development of self-tolerance to peripheral proteins during thymic selection of T-cells. Pietropaolo M, Roith Derek. (2001) Clinical cornerstone. Pathogenesis of Diabetes: Our Current Understanding. 4 (2): 1-6. (2) ..... | 12 |
| Figure 3. Primer Design for SNP rs3807830: Forward and reverse primers in bold, biotinylated primer in bold italics, sequencing primer highlighted in grey, SNP nucleotide underlined and bold. ....  | 27 |
| Figure 4. Location of the three SNPs analyzed in the study. This picture shows the three isoforms of the <i>ICA1</i> gene; it is adapted from the HapMap2 database. ....  | 28 |
| Figure 5. The Principle of Pyrosequencing (110).....  | 33 |



## **ACKNOWLEDGEMENTS**

I would first like to express my most sincere appreciation to my advisor, Dr. Massimo Pietropaolo. Thank you so much for the opportunity to work in your laboratory. I have learned so much from you and from all of the members of your lab.

I would also like to thank the other members of my thesis committee, Ms. Elizabeth Gettig, Dr. John Wilson and Dr. Michael Barmada. I really appreciate all of the time you have taken to help make my thesis the best it can be. Your encouragement and suggestions are greatly appreciated.

I would like to thank everyone who works in Dr. Pietropaolo's lab for their continued support and guidance during my time here. I would especially like to thank Dr. YiGang Chang who helped me numerous times with my project. I would like express appreciation to Dr. Steve Ringquist and the members of his laboratory who have helped me to learn and perfect the pyrosequencing method, and who have supported me each step of the way. I would like to thank Dr. Wilson, Dr. Eleanor Feingold, and Mr. David Corcoran for their help with my statistical analysis.

I would like to thank my roommates for all of their support both in and out of the classroom. I would like to thank my family and friends for all of their support along the way and for always believing in me. Lastly, to the Genetic Counseling class of 2006, thank you so much for being there for me during the past two years. Your support and friendship has meant the world to me, and I truly could not have accomplished all that I have without you.

## **1.0 INTRODUCTION**

### **1.1 PURPOSE OF THE STUDY**

Type 1 diabetes mellitus is a major public health disease. Approximately 1 million Americans currently have type 1 diabetes. Many genetic and environmental components are believed to play a role in diabetes onset; however, the majority of these factors remain unknown, or are not currently well understood. Many genetic susceptibility loci have been identified; however, few genetic factors have been identified to have strong associations with disease onset. The goal of this study is to determine if our candidate gene *ICA1* may play a role in the onset of type 1 diabetes. We will determine this potential association by analyzing single nucleotide polymorphisms within intronic regions of this gene. We will compare several SNP loci in diabetic cases and non-diabetic controls to assess whether there is a significant genotypic difference between the two populations in one or more of the SNP locations. We hypothesized that single nucleotide polymorphisms within the human *ICA1* gene are associated with type 1 diabetes mellitus and therefore that we would find a significant difference between the cases and controls at one or more SNP loci.

## **2.0 BACKGROUND AND SIGNIFICANCE**

### **2.1 DIABETES OVERVIEW**

Diabetes mellitus is a heterogeneous disorder causing dysregulation of carbohydrate metabolism. There are two major categories of this disease, termed type 1 and type 2 diabetes. Whereas type 1 diabetes results from an autoimmune destruction of pancreatic  $\beta$ -cells, the far more common type 2 diabetes is caused by the failure of pancreatic  $\beta$ -cells to meet the insulin secretory demand, which is often triggered by insulin resistance in peripheral tissue. These two forms of diabetes are inherited in a multifactorial manner. There is a third type of diabetes called MODY, (maturity-onset diabetes of the young), which is inherited in an autosomal dominant manner (1).

Type 1 diabetes (T1DM), also known as insulin-dependent diabetes mellitus (IDDM), is an autoimmune disease resulting from complex interactions between both environmental and genetic factors (2, 3). This disorder results from autoimmune destruction of the body's insulin producing cells, the pancreatic beta cells (4). Beta cell destruction is considered the hallmark feature of type 1 diabetes (5). Type 1 diabetes occurs most frequently in those of northern European decent. It is far less common in other racial groups (3, 4). Type 1 diabetes clusters in particular ethnic populations such as in Finland, with an incidence of more than 40 per 10,000 a year, and Sardinia with an incidence of greater than 30 per 10,000 a year. In the United States, about 12-14 per 100,000 individuals, ages 16 years or younger, are affected each year with type 1 diabetes. This accounts for approximately 1 million Americans (4).

A genetic predisposition to type 1 diabetes is thought to be coupled with a humoral/autoimmune response, which is believed to be caused by an environmental trigger leading to the destruction of the beta cell. A progressive loss of insulin secretion, follows, which advances the affected individual into the disease state (1, 2). The beta cells make up about 70% of the islets of Langerhans in the pancreas. Patients with type 1 diabetes have a nearly total loss of their beta

cells. However, other types of islet cells are preserved. Due to the loss of beta cells, the islets of diabetic patients are unusually small (4). Once the beta cells have been destroyed, the body can no longer effectively secrete insulin to regulate blood glucose levels, a situation which results in hyperglycemia (5). Current understanding of the environmental factors that trigger the autoimmune destruction of the beta cells in those genetically predisposed to type 1 diabetes are not well understood. Therefore, no mechanism is currently known to affectively prevent or reverse type 1 diabetes (1, 5).

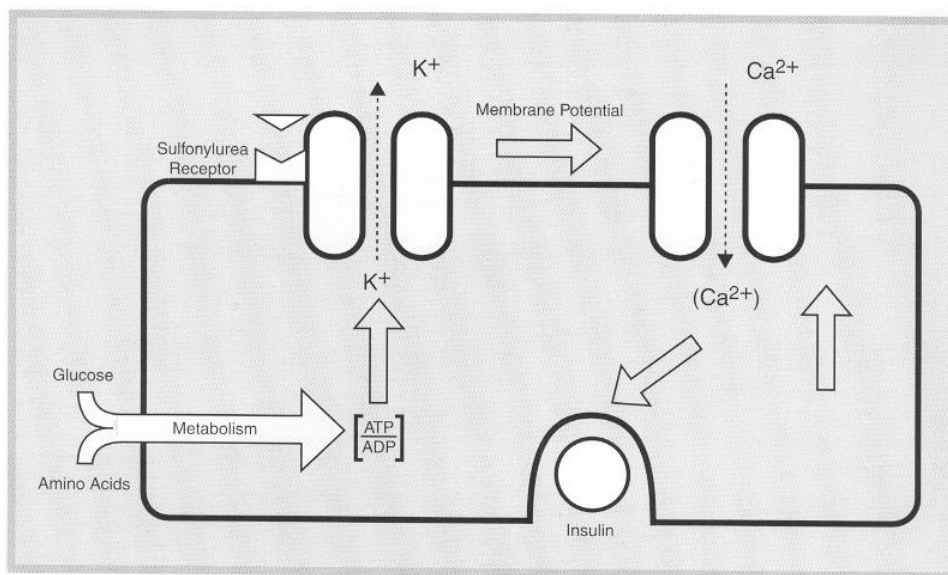
## **2.2 ENVIRONMENTAL FACTORS**

Several environmental factors have been shown to play a role in type 1 diabetes onset, specifically infection with the Rubella virus and the Coxsackie virus (6). Babies who are given cow's milk in infancy have been shown to have an increased incidence of type 1 diabetes (7, 8). Cow's milk contains bovine insulin, and it is known that humans can develop antibodies to this type of insulin (7, 8). One study showed that infants who were exposed to cow's milk during the first three months of life had higher IgG-antibodies to bovine insulin than those who were breastfed. The antibodies to bovine insulin then cross-react with human insulin (7). It is believed that the immune response to human insulin can be diverted into an aggressive immune response against the Beta-cells. This provides the environmental trigger for otherwise healthy infants to develop a beta cell specific autoimmune response (8).

Molecular mimicry is one proposed mechanism to explain how a viral infection leads to the onset of type 1 diabetes (4). In this theory, the viral particle has an amino acid sequence that is similar to the amino acid sequence of the beta cell protein glutamate decarboxylase, or GAD, which allows these two proteins to cross-react. The T-cells made by the body not only destroy the virus but also destroy the beta cells in the process of fighting the infection (4).

## 2.3 PHYSIOLOGY

Normal fasting blood glucose levels should be about 60-100mg/dL, with levels rising no higher than 140mg/dL after a meal. In fasting states, about 50% of our glucose stores are used by the brain, about 25% by the muscles, and about 25% by the intestines and other organs (2). Normally, body insulin concentration is controlled by a feedback system that is influenced by both glucose and insulin levels in the body. Insulin is also minimally responsive to amino acids and fatty acids, but it is by far the most responsive to body glucose levels (1). After food consumption, glucose is absorbed by the body causing an increase in blood glucose levels. Secreted insulin suppresses liver glucose production by inhibiting glycogenolysis. The high insulin levels in the blood then stimulate peripheral tissues to uptake glucose (2). In pancreatic beta cells metabolism of glucose drives the production of ATP. This causes the potassium channel on the beta cell membrane to close, which depolarizes the membrane. This depolarization causes a membrane potential which opens the voltage dependent calcium channel. This calcium build-up inside the beta cells leads to insulin secretion (Figure 1) (2).



**Figure 1. Glucose Induced Insulin secretion from the pancreatic beta cells. Pietropaolo M, Roith Derek. (2001) Clinical cornerstone. Pathogenesis of Diabetes: Our Current Understanding. 4 (2): 1-6.**

Insulin plays several roles in regulating blood glucose levels. In the liver, existing glucose inhibits the formation of new glucose by halting glycogenolysis as well as by stimulating glucose uptake and glycogen storage. In the fat and muscle tissue, insulin triggers glucose transport proteins (GLUT4) to uptake glucose into the cell. Glucose is metabolized in the muscles and stores are utilized for oxidation or stored as glycogen (2).

## **2.4 GENETIC SUSCEPTIBILITY**

It has been shown that multiple genes play a role in genetic susceptibility to type 1 diabetes (3, 9). Family studies have shown a clear genetic basis for the disease. However, monozygotic twin studies show a concordance rate of less than 100%, indicating that both environmental and genetic factors play a role (3). The empiric risk for monozygotic twins is 33-50% (3, 4). First-degree relatives show an empiric risk to develop type 1 diabetes of 5-7% (10-14), whereas individuals without a relative with type 1 diabetes have less than a 1% risk of developing the disorder (Table 1) (15).

The utilization of genome-wide scans has resulted in the identification of over twenty putative loci but, for now, only linkage to HLA loci appears to be unambiguous (3, 16, 17). Many of these findings have been controversial and researchers have obtained conflicting results suggesting that many of these loci are actually not linked to type 1 diabetes (12).

Two specific regions of the genome have been implicated by several researchers to play a role in the development of type 1 diabetes: the HLA region on chromosome 6p21, (*IDDM1*), and the insulin gene region on chromosome 11p15, (*IDDM2*) (15, 16, 17). Some researchers have suggested that these regions combined contribute up to 42% of genetic risk for developing type 1 diabetes (9). Others have suggested that the HLA region alone confers about 50% of disease susceptibility and the insulin gene alone contributes to about 10% of genetic susceptibility (17). Variants in the *HLA* gene, specifically in the DQ and DR regions, have been linked to type 1 diabetes (3, 9, 19).

**Table 1. Empiric risks of type 1 diabetes. These estimates are for North American Caucasian\* and Scandinavian populations\*\*, ‡ 1-15% range depending on the populations (10-15).**

|  | Empiric Risk |
|--|--------------|
| First degree relatives of T1DM probands* | 5-7%‡        |
| Individuals without relatives with T1DM* | <1%          |
| Children of affected father**            | ~6%          |
| Children of affected mother**            | ~2%          |

Studies conducted in the United States and Scandinavia have shown that the offspring of parents with type 1 diabetes have a higher risk of developing type 1 diabetes if their father is affected (6%) compared to if their mother is affected (2%) (20) (Table 1). Another group found that the recurrence risk for a child to develop type 1 diabetes if they have an affected father is 1/40, but if the mother is the affected parent their risk is 1/66 (21). Several explanations have been proposed to account for these gender differences, including genomic imprinting affecting susceptibility to T1DM manifesting as an increase in spontaneous abortions in type 1 diabetic mothers whose fetuses are genetically predisposed to develop T1DM (22,23). A second explanation includes specific maternal environmental factors that determine the fetus' level of tolerance to islet autoantigens (22), and third a preferential paternal transmission of HLA diabetogenic alleles (24).

Two single gene syndromes are known to cause autoimmune-mediated diabetes. These syndromes are termed autoimmune polyglandular syndrome type I (APS-I), also known as autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) (25,26), and X-linked polyendocrinopathy, immune dysfunction and diarrhea (XPID) (27). These two

syndromes are considered unique models and are important to understand the mechanisms involved in the loss of tolerance to self-antigens in autoimmune diabetes, as well as the associated organ-specific autoimmune disorders. In individuals with APECED, a single gene mutation can cause an array of autoimmune polyendocrine disorders, including insulin-requiring diabetes. APECED is unique among autoimmune disorders in that it is caused by mutations in a single gene pair. This autosomal recessive disease is due to homozygosity, or complex heterozygosity for a defect in the *AIRE* (autoimmune regulator) gene, which is localized on chromosome 21q22.3. Over 31 different mutations within the *AIRE* locus have been described, the most frequent a 13-bp deletion in exon 8 (1085-1097 deletion). This specific mutation appears to have originated in the Norwegian population. (28). No specific genotype-phenotype correlations have been identified at the *AIRE* locus (28).

Currently it is believed that the *AIRE* gene codes for a transcription factor (29). The nature of the immune abnormalities still remains unclear even though dysregulation of T-cell function as well as cytokine patterns have been implicated in the pathogenesis of this polyendocrine disorder. It is important to note that individuals affected with APECED develop antibodies towards self- molecules such as glutamic acid decarboxylase (GAD) and islet cell antibodies (ICA). Individuals with type 1 diabetes develop antibodies to these same self-molecules.

XPID is an X-linked recessive immunologic disorder characterized by early onset of type 1A diabetes mellitus, autoimmune thyroiditis, autoimmune enteropathy, hemolytic anemia, atopic dermatitis including food allergies, and immunodeficiency leading to fatal infections. This syndrome is also referred to as X-linked autoimmunity-immunodeficiency syndrome (XLAAD). A number of mutations within the *JM2* gene, which is localized in the peri-centromeric region of the X chromosome (Xp11.23-Xq13.3), have been documented in patients affected by XPID (27).



## 2.5 ROLE OF THE HLA COMPLEX

The Major Histocompatibility Complex, (MHC), is located on chromosome 6p. Several of the principle genes in this region encode for human leukocyte antigens, (HLA). Two separate classes of HLA complexes exist. These two complexes are cellular surface glycoproteins, which differ in structure and function, as well as in tissue location (9).

The class I HLA molecules is composed of a heterodimer composed of an  $\alpha$ -chain, which is the heavy chain, and a  $\beta$ 2-microglobulin. The HLA I molecule is anchored into the cell membrane by the heavy chain. The chain contains an extracellular hydrophilic region, a transmembrane hydrophobic region, and an intracytoplasmic hydrophilic region. The extracellular region is composed of three alpha subunits which comprise the antigen-binding region of the HLA I complex (9).

The class II HLA molecules are composed of an  $\alpha$  and a  $\beta$  glycoprotein chain. In the class II HLA molecule, both of these chains are anchored in the membrane. Each HLA II chain is also composed of an extracellular, transmembrane and intracellular region. The extracellular domain in each class II HLA molecule contains its own alpha and beta subunits. The  $\alpha_1$  and  $\beta_1$  subunits together make up the antigen-binding region (9).

The genes that make up the class I MHC molecules include *HLA-A*, *HLA-B*, and *HLA-C*. The class II MHC molecule is composed of the genes *DR*, *DQ* and *DP*. Two other genes clustering in this region that are important for processing antigens include the *TAP* and *LMP* genes. The third region of the MHC complex, MHC class III, codes for several genes that have a variety of functions.

Class I MHC molecules are expressed in most nucleated cells; however, class II MHC molecules are only expressed in selected tissues, including B lymphocytes, dendritic cells, macrophages, and activated T lymphocytes. Both MHC class I and class II molecules play a role in antigen T-cell presentation (9). Class I antigens are primarily recognized by CD8+ cells, (cytotoxic T cells), whereas class II antigens are primarily recognized by CD4+, (helper T cells). Amino acid changes in the HLA-DQB-chain have been shown to influence the interaction between the class II molecule and the presentation of the antigen to the T-cell receptor. This interaction alters the specificity of the immune response to both foreign and self antigens. It appears that amino acid changes in the DQ $\beta$ -chain as well as the DQ $\alpha$ -chain are involved in an

individual's susceptibility to type 1 diabetes. Genetic susceptibility has been strongly associated with the HLA haplotypes-DQA1\*0501, DQB1\*0201 and DQA1\*0301, DQB1\*0302 (9). This finding has been supported by recently published data by Wen *et al.*, which demonstrates the contribution of the HLA-DQ molecules to the development of autoimmune-related diabetes in a mouse model (29). Using this “humanized” animal model the authors demonstrated that by replacing *HLA-DQ8* (believed to play a role in diabetes susceptibility) with *HLA-DQ6* (believed to be a protective allele) diabetes was prevented (30).

Studies in the non-obese diabetic (NOD) mouse have also demonstrated the importance of the class II DQ molecules. In these studies, amino acid alterations in the I-A $\beta$  chain, which is equivalent to the human class II *DQB1* locus, can prevent mice from developing diabetes (31,32). It has also been shown in the NOD mouse model that genetic alterations in the *I-E* gene, which is equivalent to human *HLA-DR*, appears to confer diabetes resistance (31,33). One group also demonstrated that by treating the NOD mouse with a monoclonal antibody that interacted with the animal's class II molecule, progression to type 1 diabetes could be prevented (34). These studies support the role of *HLA-DQ* and *HLA-DR* in development of type 1 diabetes.

The mechanisms by which the class II genes influence both diabetes susceptibility and protection are still unclear. The crystallized HLA class II molecule structure was characterized in 1993 by Brown *et al* (35). One hypothesis for the altered susceptibility is that effective antigen binding is dependent on the specific conformation of the DQ antigen binding site. It has been postulated that an amino acid change at certain positions of the DQ molecule will lead to a conformational change that affects antigen-binding, causing an alteration in the affinity of the class II molecule and the “diabetogenic” peptide(s). The three-dimensional structure of DQ8 complexed with the insulin B chain has been determined by Lee *et al.* using X-ray crystallography. This binding represents an important T-cell epitope for the islet-infiltrating T-cells in NOD mice (36).

There is still debate as to whether the DR molecule plays an important role in an individual's susceptibility to type 1 diabetes (37). DR alleles have been shown to be in linkage disequilibrium with DQ alleles, and some investigators have argued that the DR alleles are not the primary factor conferring disease susceptibility. However, others believe that the DR locus plays its own independent role in type 1 diabetes susceptibility (37).

In individuals with type 1 diabetes, the prevalence of HLA-DR2 is decreased. The DQA1\*0102/DQB1\*0502/DRB1\*1601 appears to account for the majority of disease susceptibility in DR2-associated cases of type 1 diabetes (38). It is interesting to note that, in many populations the frequency of the DQB1\*0602 allele is rarely found in individuals with type 1 diabetes. This suggests that this allele plays a protective role against diabetes onset. Currently, this seemingly protective DQB1\*0602 allele is considered a clinical trial exclusion criterion, preventing first degree relatives with this allele from participating in clinical trials such as the Diabetes Prevention Trial 1 (DPT-1) (39), and the Type 1 Diabetes TrialNet. Both of these clinical trials have been designed to delay and ultimately prevent the progression to type 1 diabetes in individuals who are considered to be at high risk for disease development.

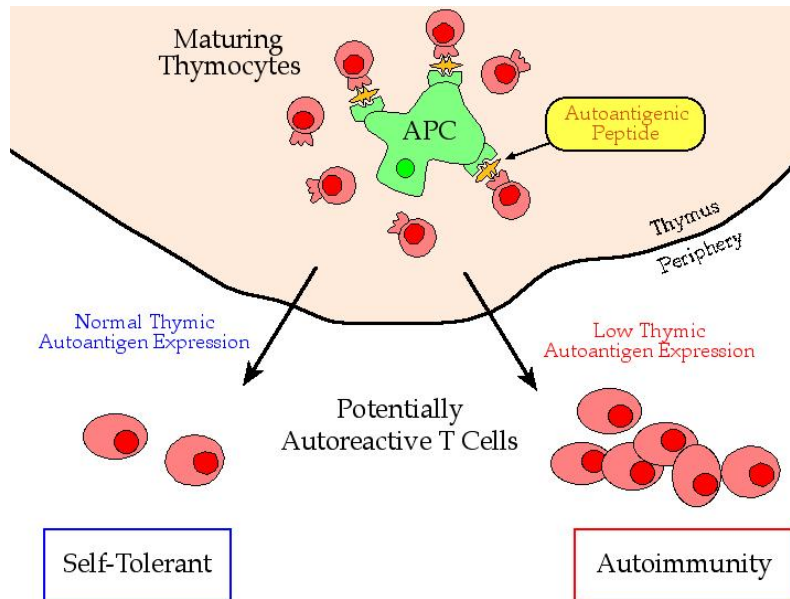
There are several proposed mechanisms that might explain the autoimmunity associated with type 1 diabetes and of the role of the HLA complex in disease pathogenesis (9). The first hypothesis is that the self-antigen (autoantigen) is presented to the antigen-presenting cell (APC) through the same mechanism that a viral peptide is presented. The body then generates an immune response directed against the antigenic peptide. However, in individuals who develop type 1 diabetes, the antigen is actually a self-peptide which leads instead to an autoimmune response.

The second hypothesis is that the alpha and beta chains of the T-cell receptors located on the outside of the HLA-peptide can interact with foreign molecules. These foreign molecules, specifically bacterial or viral particles, may induce an immune response against self-antigens by activating auto-reactive T-cell clones.

The third hypothesis is that the cell loses the ability to transport antigenic peptides from the cytoplasm to the endoplasmic reticulum because of a defect in the gene product for the antigenic peptide transporter. This would cause a decreased amount of peptides to bind to class I receptors causing a surplus of empty class I molecules. These empty receptors may then bind to viral or bacterial or self peptides that they would not normally bind to in optimal conditions. These new peptides may then lead to an autoimmune response.

## 2.6 ROLE OF THE INSULIN GENE

The insulin gene, found at 11p15, is another locus that has shown linkage to type 1 diabetes. The insulin gene (INS) is linked to the VNTR (variable number of tandem repeats) region. The VNTR alleles are divided into two groups; class I is the shorter class and has 26-63 repeats, whereas class III has 140-200 repeats. It is believed that allelic variation in this locus alters the levels of insulin mRNA in the pancreas (40). Class III VNTRs are protective against diabetes, while class I VNTRs confer a 2-5 fold increase in risk. (40, 41). Inheriting one class III VNTR allele decreases the risk of developing type 1 diabetes by 60%. The protective class III VNTR- linked *INS* genes have lower expression levels of insulin mRNA in the pancreas than the Class I VNTR-linked insulin genes (41). However, protective class III VNTRs are associated with higher levels of INS mRNA expression in the thymus. This high thymic expression may promote negative selection of the T-lymphocytes. (40). This theory is based on the observation that the body's self-recognition occurs in the thymic medulla. It is here that antigen presenting cells (APC) present self-peptides to the T-cells to allow the body to recognize these molecules as 'self'. This leads to apoptosis of the T-cells specific to this self-antigen and/or to the loss of the specific receptor for this self antigen on the T- cell. It is this 'self' versus 'non self' distinction that allows our body to recognize its own peptides so that it does not attack them at later times (the autoimmune response) or that leads to an immune response directed against foreign molecules. The APC presents peptides to both the Class I and class II MHC molecules (42). In this negative selection model, the higher levels of class III VNTR insulin presentation in the thymus, where T-cell self recognition occurs and promotes greater recognition of self-peptides. This recognition leads to the destruction of self-reactive T-cells. If the T-cells which degrade insulin are destroyed, the autoimmune response is halted, thereby decreasing disease incidence. The class I VNTR has decreased insulin expression in the human thymus. This does not allow self-recognition to occur or the destruction of the T-cells specific to this self-antigen. Therefore, the body may still have autoreactive T-cells for the insulin protein (Figure 2) (2).



**Figure 2. General hypothesis for the development of self-tolerance to peripheral proteins during thymic selection of T-cells. Pietropaolo M, Roith Derek. (2001) Clinical cornerstone. Pathogenesis of Diabetes: Our Current Understanding. 4 (2): 1-6. (2)**

## 2.7 IMMUNE SYSTEM

Diabetic autoimmunity is primarily mediated by T-lymphocytes. Pre-diabetes, sometimes called insulinitis, occurs when a mixed population of leukocytes (white blood cells) invade the islets of Langerhans. This stage can last for years in humans (5). The primary types of leukocyte invading the islets are the T-cells. Studies in the non-obese diabetic (NOD) mouse have shown that diabetic mice do not develop diabetes if they are athymic (lacking a thymus) or lymphopenic (having few T-lymphocyte cells). It is known that the thymus consists mainly of lymphatic tissue and is the major site for T- cell differentiation. Studies have also demonstrated that type 1 diabetes can be transferred by injecting T- cells from diseased NOD donors into healthy NOD recipients. Both CD4<sup>+</sup> cells and CD8<sup>+</sup> cells have been shown to play a role in the pathology of diabetes. In support of the role of islet cell autoimmunity in type 1 diabetes, it has been thoroughly documented that individuals with type 1 diabetes have a mononuclear cell infiltration in the pancreatic islets, with predominantly CD4<sup>+</sup> and CD8<sup>+</sup> cells, which are likely involved in

selective pancreatic  $\beta$ -cell destruction. These cells respond to several antigens that are synthesized in the beta cells. As previously mentioned, the antigens for the CD4<sup>+</sup> T cells are presented by the MHC class II molecules located on specialized antigen presenting cells. The peptide targets for the CD8<sup>+</sup> T cells are displayed by the MHC class I molecules present on most cell types (5).

Self-tolerance refers to a set of events that very specifically eliminates or anergizes a potentially autoreactive T-cell, which ultimately induces a “tolerance” to self molecules. Self-tolerance is established and maintained through complex mechanisms that occur in both the thymus, which is responsible for central tolerance, and the peripheral lymphoid organs, which are responsible for peripheral tolerance. It is believed that proteins with tissue-restricted or peripheral expression are not available for thymus presentation. Therefore, self-tolerance for these proteins must be achieved through mechanisms of peripheral tolerance. One hypothesis for the lack of self-tolerance that occurs in type 1 diabetes involves either a failure of negative selection of autoreactive T-cells occurring in either the thymus or in the periphery, or a breakdown in tolerance to B-cell-specific antigens (Figure 2). This theory has been supported by studies that have demonstrated prevention of diabetes in the NOD mouse and BB (diabetic) rat models following thymic transplantation of islet antigens or expression of putative islet cell autoantigens (39,43). Recent literature suggests that molecules appearing to have tissue-restricted expression may also be expressed in the thymus (39). It is known that genes encoding several type I diabetes related autoantigens, including insulin, IA-2, GAD, and the neuroendocrine antigen ICA69, are transcribed in the human thymus during the fetal period as well as in early childhood. Thymic human *insulin* gene transcription has been reported by several research groups, and similarly insulin, glucagon, GAD and ICA69 transcripts have been detected in both the rat and mouse thymus (39,45,48). Transcripts of several self-molecules have also been detected in the thymus including pancreatic and thyroid hormones, neuroendocrine molecules, and peripheral self-proteins. This raises the possibility that expression of self-antigens in the thymus may be imperative for the development of self-tolerance.

Further research in mouse models has shown that the thymus harbors cells called peripheral-antigen-expressing or PAE cells. These PAE cells express peripheral antigens such as insulin. Transgenic mice studies have provided evidence to suggest that these PAE cells have an important role in self tolerance (48). Similar cells have recently been described in humans in

both the thymus and peripheral lymphoid tissues. These cells were found to express islet cell antigens such as insulin/proinsulin, GAD and IA-2. Further phenotypic characterization of PAE cells in both the mouse thymus and the human lymphoid tissues indicates that they express phenotypic markers of antigen presenting cells. Given their phenotype and their role in self-antigen expression, they have been referred to as self-antigen presenting cells (45).

T-cells express functional T-cell receptors (TCR) on their cell surface. Once this receptor has been expressed, it is subject to either positive or negative selection (46,47). These selection events are dependent on the interactions between the TCR, the MHC molecule, and the antigenic self-peptide. Positive selection occurs when the thymic stromal cells, bearing the MHC molecules and containing a self-peptide fragment, direct the T-cell receptors on the developing thymocytes to continue maturation into functionally mature T-cells. T-cells with receptors that are unable to bind to the MHC molecule with sufficient affinity are not driven to mature and expand, and eventually these cells die. Negative selection is a process that is not currently well understood. This refers to the T-cell elimination or anergy which induces self-tolerance. During negative selection it is believed that factors such as antigen load, and affinity for self-antigens influence the result and determine whether the T-cell will undergo cell death or clonal anergy. Due to the unique immune system qualities, including the random generation of TCRs in the initial thymocyte as well as the process of thymic selection, the peripheral T-cell collection is unique in each individual, even in identical twins.

Autoimmunity is believed to be caused by an imbalance between tolerance induction and immune responsiveness. Both of these processes are dependent on the class I and class II MHC molecules and their ability to present critical antigenic peptides. It is believed that in genetically susceptible individuals, certain class II molecules may insufficiently present self-peptides, leading to inadequate negative T-cell selection. This inadequate negative selection could fail to eliminate a T-cell population that could later become activated against self-molecules thereby creating an autoimmune response. Researchers Nepom and Kwock have described the association of the HLA-DQ alleles and type 1 diabetes using this molecular mechanism (49).

Further immunologic studies have suggested other mechanisms which appear to play an important role in type 1 diabetes. Pancreatic dendritic immune cells (DCs) transfers have been shown to prevent diabetes in NOD mice (50-52). The role that DC play in type 1 diabetes is not

currently understood, but it is suggested that patients with type 1 diabetes may have fewer DCs, or the DCs they have may be functionally impaired.

Further research has explored the association between type 1 diabetes and transplant tolerance. Data obtained in mouse models suggests that the diabetic phenotype and the resistance to transplantation tolerance may co-exist; however, these two features may not be under the same genetic control, and may in fact have different genetic causes (53). If these two mechanisms do have distinct genetic etiologies, then the strategies developed to treat autoimmune diabetes and those developed to prevent recurrent autoimmunity should be approached differently from those instilled to induce a transplant tolerance.

It has also been suggested that alterations in the function and number of several regulatory cells may play a role in the generation of an autoimmune state in type 1 diabetes (54,55). Dysfunction and/or loss of CD1-restricted T-cell, T cells with  $\gamma/\delta$  receptors,  $CD4^+/CD25^+$  T cells and NKT cells have been proposed to contribute to diabetes pathogenesis via insufficient suppression of autoreactive T-cells. One example of this is represented by discordant twin studies. These studies show that the levels of CD-1 restricted T-cells appear to be diminished in the affected twin relative to the non-affected twin (55). The antigens responsible for antigen T-cell activation are currently unknown, as are the mechanisms by which these cells exert their effect on the immune system.

## **2.8 AUTOANTIBODIES**

The T-cell build up in the beta cells of pre-clinical diabetics ultimately leads to an accumulation of antibodies to a variety of islet cell proteins. However, the role that these autoantibodies play in beta cell destruction is not currently well understood. Autoantibodies have been shown to develop against insulin, glutamic acid decarboxylase (GAD), protein tyrosine phosphatase-2 (IA-1), and to islet cell antigens (ICA) (3). In the general population antibodies to these proteins are fairly low, existing in only about 2-4% of people (3). However, antibodies to these proteins are commonly found in newly diagnosed diabetic patients as well as in their relatives who later go on to develop clinical diabetes. Studies by Verge *et al.* (56), have indicated that in relatives of diabetics who later go on to develop the disease, 98% have one or more antibodies to insulin,



GAD and/or the specific islet cell autoantigen ICA512bdcAAs. About 80% have antibodies to two or more of these proteins. It has also been found that these antibodies serve as good predictors among relatives of diabetic patients (56). Currently the most widely utilized screening strategy used to identify individuals who are at risk to develop type 1 diabetes (generally deemed so due to an affected first degree relative), is to screen for autoantibodies to the islet antigens, GAD65 and IA-2. If one of these screening tests yields a positive result, the results are confirmed using an additional serum sample. Once this conformation has been established, both insulin autoantibodies and ICA autoantibodies are measured and the first phase (1+3) insulin response to bolus intravenous glucose (AIRg) is determined. By combining both of these immunologic and metabolism studies, it is believed type 1 diabetes progression can be predicted with 50% accuracy within 5 years of follow-up, and 80-90% accuracy within 10 years of follow-up. (57, 58) Genetic typing to assess the HLA alleles for a susceptibility or protective affect is also performed. Although HLA typing alone is not adequate to predict diabetes onset, it is believed that the evaluation of the HLA genes in first degree relatives of type 1 diabetic patients does prove to be useful. (57).

## **2.9 CANDIDATE GENE APPROACH**

The HLA region on chromosome 6p21, (the IDDM1 region), and the insulin gene on chromosome 11p15, (the IDDM2 region), account for approximately 42% and 10% of an individuals risk to type 1 develop diabetes, respectively. Due to the fact that these two genetic components do not contribute 100% to genetic susceptibility, it is important to try to discover the additional factors, specifically other genes, which play a role in disease susceptibility. In addition to genome-wide scans and positional cloning, which proved to be less successful than expected, the candidate gene approach has been proposed to discover novel type 1 diabetes susceptibility loci.

The candidate gene approach takes factors into account other than chromosome location, which is the only consideration in positional cloning. This approach, like positional cloning, looks at an area of the genome that is associated with disease, in our case type 1 diabetes. However, it also looks for a gene associated with the disease phenotype that lies in or near the

linked loci. (58, 59). This approach has been used to discover the genes for diseases such as Retinitis Pigmentosa, HNPCC and Waardenburg Syndrome (58, 59) and is currently being utilized to study other multifactorial conditions such hypertension, various types of cancer, and diabetes (85).

Due to the fact that the candidate gene approach can take into account both location and function, we have adopted this approach for our current study. The candidate-gene *ICA1*, which is the focus of our study, was selected because of an *a priori* hypothesis suggesting the etiological role of the *ICA1* protein product, ICA69, in diabetes pathogenesis in both human and NOD mouse animal models (16, 45, 60-66). Several studies by our group and others have suggested that the *ICA1* gene may have an important biological role in the genetic regulation of ICA69 expression in the thymus (16, 67-72). It is also interesting to note that ICA69 expression has been reported in thymic dendritic cells (45). Recent data suggests that there is reduced ICA69 expression in the thymus of the NOD mouse compared to diabetes resistant mice (C57BL/6, B6, NOD.B10-H2<sup>b</sup>, ALR), and that this may be associated with an increased predisposition to type 1 diabetes (16). This implies that high levels of ICA69 expression in the thymus may be important for clonal deletion of ICA69 reactive lymphocytes (45). Allelic variations in the regulatory sequences for *ICA1* could influence the expression of ICA69 in antigen presenting cells, such as DCs, during thymic development (45, 73-78). These variations may cause a defect in the transcriptional activity of ICA69 in the thymus (67).

Based on this information, we hypothesize that higher levels of ICA69 protein in the thymus promote negative selection of ICA69-specific T lymphocytes. A defect in the transcriptional activity of *ICA1* may cause failure to induce ICA69 tolerance during thymic development. It is possible that polymorphic regions within promoter elements of *ICA1* may exert a regulatory effect by strengthening nucleosome formation, thereby altering the local chromatin structure and consequently decreasing the efficiency of transcription, a mechanism which has been previously described for the *TIDM2* gene (39, 79-82). Therefore, a transcriptional defect of *ICA1* in the thymus may play a critical role in the autoimmune destruction of the pancreatic islet cells by autoreactive T-lymphocytes ultimately leading to type 1 diabetes.

## 2.10 ISLET CELL AUTOANTIGEN 69KDA (ICA69)

The candidate gene we have selected for our study is *Islet Cell Autoantigen 1*, which codes for the protein Islet Cell Autoantigen 69kD (ICA69). This protein is one of the specific protein targets for autoimmune response that results in diabetes (85,86). About 25% of recent onset type 1 diabetic patients have antibodies to this protein (89). First degree relatives of diabetic patients, who later develop disease, also have antibodies that react to ICA69 (87). This protein has been shown to play a role in the humoral and/or cellular immune response (87). Pietropaolo *et al.* first identified this protein by screening the human islet cDNA expression library  $\lambda$ g11 with the sera from the pre-diabetic relatives of affected individuals. A number of these individuals were ICA69 antibody positive and subsequently progressed to overt type 1 diabetes. The ICA69 antigen was specifically detected in beta cell lines (87).

The ICA1 gene is found on chromosome 7p22. There has been research to show that type 1 diabetes is linked to this area of the genome (89, 90). It was originally believed that this linkage was due to the glucokinase gene, which is located at 7p15.3-p15.1. Glucokinase is believed to be a glucose sensor. It also has the ability to generate metabolic signals in the pancreas. This enzyme is important for regulation of insulin secretion during glycolysis (91). The role of this enzyme in insulin secretion, as well as its location near an area of the genome with strong linkage to diabetes, made it a prime candidate gene for type 1 diabetes. It was originally believed that a variant in the promoter for the glucokinase gene was responsible for diabetic susceptibility. However, work by Lotfi *et al.* suggests that this variant is not associated with type 1 diabetes (91). This research raises the question that it may not be the glucokinase gene that is causing linkage to the 7p loci in diabetes, but that another gene located on 7p, such as our candidate gene *ICA1*, may in fact be the link between this locus and type 1 diabetes (91).

ICA1 is made up of 14 exons, which are 39-271 bp in length, and three 5' untranslated regions. This gene is thought to highly conserved both in human, rat, and mouse homologues as well as in exon/intron partitioning (85, 86). Mouse and human intron-exon junctions are identical with introns ranging in size from 94 bp to 24 kb (86). ICA69 mRNA has been detected primarily in endocrine cells comprising the islets (87). This protein has also been shown to participate in the neurosecretory process in its *C. elegans* homologue (92). This may suggest that ICA69 is involved in the insulin secretory pathway in the beta cells (85).

Further evidence to suggest that ICA69 plays a role in diabetes was found by Roep *et al.*, who demonstrated T-cell reactivity to ICA69 (64). This autoimmune T-cell response against self ICA69 was found to be higher in recently diagnosed type 1 diabetics compared to those who had had been living with disease (88). T-cell reactivity to ICA69 has also been seen in the NOD mouse model (87). Work in the NOD mouse by Winer *et al.* shows that injection of T- cells specific for Tep69 (dominant NOD T-cell epitope in ICA69) may play a role in the acceleration of islet destruction in the NOD mouse (93). The research by Winer's group suggests that injection of T-cells specific for this peptide causes an acceleration of the diabetic autoimmune response. Studies also show that neonatal injection of Tep69 into mice leads to a reduced incidence of type 1 diabetes, (94), suggesting that ICA69 may be important in disease prevention.

Research has shown that the environmental trigger initiated by cow's milk containing bovine serum albumin may also play a role in the autoimmune targeting of ICA69. BSA contains p69, which is a protein identical to ICA69 (7). Structural homologies exist between the ABBOS region of bovine serum albumin and the Tep69. (87, 95). BSA has been shown to trigger T-cell proliferation against these p69 molecules (7). These findings suggest the possible mechanism for cow's milk to act as an environmental trigger for the onset of type 1 diabetes (7). This model is based on antigen molecular mimicry between the antigen for cow's milk and ICA69 protein (86).

As previously mentioned, studies comparing the non-obese diabetic mouse (NOD) to the B6 non-diabetic mouse show that ICA69 has significantly reduced expression in diabetic mouse thymic tissue compared to the thymic tissue of its non-diabetic counterpart. Real time PCR studies indicate that ICA69 may not be expressed at all in the thymus of the NOD mouse (41). This peptide expression in the thymus is an important part of the immune response in terms of T-lymphocyte selection and maturation (Figure 2). In the NOD mouse, a direct link between impaired thymic negative selection and autoimmunity have been reported (95). The decreased expression of ICA69 in the thymus may lead to the decreased destruction of self-reactive T-cells which induce the autoimmune response responsible for beta cell destruction (41).

Unpublished data by Chang and Pietropaolo *et al.* suggests that reduced expression of ICA69 in the mouse thymus may be due to a single nucleotide polymorphism (SNP) in the promoter region of the *Ical* gene in mouse (equivalent to the *ICA1* gene in humans). Pietropaolo's group originally discovered, using luciferase assays, that the NOD *Ical* putative

promoter element had a 4-fold reduction in transcriptional activity when compared with the wild type B6 mouse putative promoter. Further studies showed that a single SNP within *Ical* gene was associated with a substantial reduction of transcriptional activity. By applying site-directed mutagenesis in NIT-1 cells (cell line derived from NOD mice), the Pietropaolo group demonstrated that the mutation of the NOD SNP to its wild type base significantly reestablished transcriptional activity in the transfected cells. It was also shown that by performing site directed mutagenesis in the wild type cells and inserting the NOD SNP into the promoter region, transcriptional activity was significantly reduced. EMSA assays were also performed demonstrating that the oligonucleotide containing the NOD SNP of interest had markedly decreased nuclear protein binding, which further supports their results. By applying the computational analysis program designed by Corcoran *et al.* it was discovered that this SNP was located at an Egr-1 binding site (96).

Past research suggests that *ICAI* is an important candidate gene in type 1 diabetes. This evidence also suggests that SNPs within this gene may play an important role in progression to type 1 diabetes. At the initiation of our study, all validated SNPs within the *ICAI* gene were intronic, and no exonic SNPs had been identified. Also at the beginning of our study, no human promoter SNPs had been identified. Because current data suggests that SNPs within the mouse promoter region cause a functional difference in gene transcription, we elected to analyze the non-coding, (intronic), regions of the human *ICAI* gene.

## **2.11 THE PYROSEQUENCING METHOD**

The information obtained from the Human Genome Project provided researchers with the opportunity to study the risk factors for human diseases at the genotype level. As a result, DNA sequencing is one of the most important features of science today. Following the publication of the Human Genome Project, there was an immediate need for a new robust sequencing technique that was able to produce a high data output to compliment the older Sanger method (97). Many researchers pursued this goal and attempted to develop a new and more efficient technique (98, 99). The need to sequence single nucleotide polymorphisms as well as express sequence tags caused an even greater need for a new genotyping method that was both simple and accurate.

The idea of sequencing-by-synthesis was first proposed by Hyman in 1988 (99). Hyman's sequencing technique allowed for the detection of nucleotide incorporation using a primer-directed polymerase extension. Other researchers have developed protocols based on the use of fluorescently labeled nucleotides (60, 99-105). However Metzker *et al.* showed that by using this method, there was poor incorporation of these fluorescently labeled nucleotides, which increased the need for a more robust sequencing technique.

More recently, Ronaghi *et al.* proposed yet another technique known as pyrosequencing (106). This technique utilizes natural nucleotides, rather than their fluorescent counterparts. The basis of this method is the utilization of the pyrophosphate that is released during the DNA polymerase reaction. This pyrophosphate is converted into ATP, which can be transmitted into visible light. This technique also had limitations. The template DNA strand had to be washed thoroughly between nucleotide additions, which was very time consuming. In addition, the templates that were not bound to a solid support were more difficult to sequence (99). In 1998, Ronaghi *et al.* proposed a solution to the wash problem, the incorporation of a fourth enzyme into the pyrosequencing process. This enzyme, known as Apyrase, degrades the nucleotides during the pyrosequencing process after the addition of each nucleotide. Ronaghi's updated pyrosequencing technique was used in our current research (103).

The pyrosequencing technique has proven to be successful for a variety of sequencing projects. It has been shown to be useful in both confirmatory sequencing as well as *de novo* sequencing. It has also been used for the re-sequencing of diseased genes and for the sequence determination of difficult secondary DNA structure (103). This technique is performed in a 96 well plate format using laboratory automation (107). Theoretically more than 5000 samples can be analyzed in 8 hours (103). In addition, the technology is quite sensitive and requires minimal PCR product for each individual sequencing reaction (107). The technology is rapid: When sequencing a small area of approximately 10 nucleotides total per sample, the sequencing of an entire 96 well plate can take only 10-20 min. This technique also allows for the detection of haplotypes by analyzing SNPs that are in close proximity to each other (108).

Pyrosequencing provides high quality sequencing for short stretches of DNA, and it can resolve heterozygous nucleotides by out-of phase sequencing (107), making the sequencing method very effective for SNP genotyping. The technique allows for the sequencing of all possible SNP variations in a single well or tube. The pyrogram readouts, (the graph data

produced by this method, see materials and methods for further detail), provide a clear distinction between the various genotypes and allele combinations and can differentiate between homozygous and heterozygous individuals. The automated system for liquid-phase pyrosequencing, called the PSQ 96 system, can score the raw SNP data. This allows the computer to determine whether the run SNP run was successful. The computer process removes the need for human sequence editing (107).

One study which analyzed 26 SNPs from 1600 DNA samples found that this technique assigned the expected genotype in 99.4% of the cases. The fact that this output was not 100% was attributed to insufficient signal or noise caused by low efficiency in the prior PCR amplification step (104).

## 2.12 SPECIFIC AIMS

**Specific Aim 1:** To determine the frequencies and genotypes of several single nucleotide polymorphisms within intronic regions in the *ICAI* gene.

**Specific Aim 2:** To determine the association between SNPs in the *ICAI* gene and type 1 diabetes.

**Hypothesis:** Single Nucleotide Polymorphisms within the *ICAI* gene differ between cases and controls and play a role in the onset of type 1 diabetes.

**Plan:** We will genotype diabetic case and control subjects by first using polymerase chain reaction to amplify the area of interest around each SNP and then by using the pyrosequencing technique to obtain genotype data for each sample. We will then use appropriate statistical analysis to determine if there is a statistical difference between case and control subjects.

### **3.0 MATERIALS AND METHODS**

#### **3.1 STUDY POPULATION**

The population in this study includes N=79 control samples and N= 73 diabetic samples (termed “original diabetic samples”). All samples were collected in the Allegheny County region of Pennsylvania. The 73 cases were randomly selected from a roster of patients of the Children’s Hospital of Pittsburgh whose T1DM was diagnosed according to standard National Diabetes Data Group Criteria by careful assessment of Dr. Dorothy Becker and her team (109). Ninety five percent were Caucasian and their age ranged between 5 and 40 years. Some of these newly diagnosed T1DM patients were HLA-DQ typed in Dr. Trucco’s laboratory.

A second set of N= 33 diabetic samples (termed “additional diabetic samples”) were obtained from the Children’s Hospital of Pittsburgh study. This study includes approximately 1000 type 1 diabetic patients; it also included serum samples from 7,000 first degree relatives of type 1 diabetic probands, of whom 130 had progressed to insulin dependent diabetes. The 33 samples we obtained from this study included a random sample of individuals who were recently diagnosed with type 1 diabetes.

The 79 controls consist of randomly selected healthy Caucasian individuals from Allegheny County, Pennsylvania (age range 5-40 years) who were blood donors at the Pittsburgh Blood Bank.



### 3.2 SNP IDENTIFICATION

A total of intronic 44 SNPs had been identified in the *ICA1* gene at the start of our experiment. To ensure that the SNPs we analyzed were ‘true’ sequence variants, (variants that actually contain two different nucleotide possibilities at each SNP location), SNPs that had been previously validated were favored, or if not previously validated, SNPs were favored if they were present in multiple data bases. A total of 24 ‘true’ intronic SNPs were identified using these criteria. Later, two additional ‘true’ SNPs were added for a total of 26 SNPS. The *ICA1* SNPs considered in this study are listed in Table 2. The *ICA1* SNPs which have been identified but were excluded from our current study because they were present only in a single data base and were not previously validated are listed in Table 3. SNP locations were identified using a blast search for sequence homology using the NCBI blastn command (<http://www.ncbi.nlm.nih.gov/BLAST/>). *ICA1* locations were identified in the *Homo sapiens* BAC RP11-560C1 clone which includes the complete sequence from 7p22-p21 (Table 4).

**Table 2. *ICA1* SNPs considered to be ‘true’ which were previously validated or found in multiple databases.**

|           |            |           |           |            |
|-----------|------------|-----------|-----------|------------|
| rs6463765 | rs2110332  | rs874721  | rs3807824 | rs1008038  |
| rs3807852 | rs3757524  | rs3823833 | rs3757515 | rs759262   |
| rs3807849 | rs2348895  | rs3807830 | rs1981715 | rs10276256 |
| rs3823838 | rs3757522  | rs3823832 | rs3807818 | rs9472     |
| rs887848  | rs2058519  | rs3757510 | rs3807816 |            |
| rs2058278 | rs10264092 |           |           |            |

**Table 3. *ICA1* SNPs which were not used in this study because they were not previously validated and were identified in only a single database.**

|           |            |           |           |           |
|-----------|------------|-----------|-----------|-----------|
| rs1860532 | rs4720763  | rs2058522 | rs2192346 | rs887850  |
| rs2110333 | rs3807839  | rs3757523 | rs6463777 | rs6463778 |
| rs3807827 | rs10952094 | rs1861035 | rs4236407 | rs3779363 |
| rs3779360 | rs3807809  | rs3779357 | rs3757509 | rs1035096 |

**Table 4. SNP locations within the ICA1 gene.**

|    | <b>SNP</b> | <b>Region</b> | <b>SNP location in the <i>ICA1</i> gene (located within the designated base pairs)</b> |
|----|------------|---------------|--|
| 1  | rs3807852  | Intronic      | 175117-175167  |
| 2  | rs3823838  | Intronic      | 6651-6683  |
| 3  | rs3757524  | Intronic      | 30204-30251  |
| 4  | rs3757522  | Intronic      | 40704-40749  |
| 5  | rs874721   | Intronic      | 58827-58863  |
| 6  | rs3823833  | Intronic      | 63176-63226  |
| 7  | rs3807830  | Intronic      | 70244-70288  |
| 8  | rs3823832  | Intronic      | 74304-74339  |
| 9  | rs3807824  | Intronic      | 80924-80953  |
| 10 | rs3757515  | Intronic      | 86642-86692  |
| 11 | rs3807816  | Intronic      | 115545-115595  |
| 12 | rs1008038  | Intronic      | 126684-126733  |
| 13 | rs2058278  | Intronic      | 152563-152505  |
| 14 | 6463765    | Intronic      | 162883-162850  |
| 15 | rs887848   | Intronic      | 20473-20435  |
| 16 | rs2110332  | Intronic      | 25593-25548  |
| 17 | rs2348895  | Intronic      | 40026-39988  |
| 18 | rs2058519  | Intronic      | 55209-55161  |
| 19 | rs1981715  | Intronic      | 90295-90250  |
| 20 | rs3807818  | Intronic      | 99723-99675  |
| 21 | rs759262   | Intronic      | 1841-1793  |
| 22 | rs3807849  | Intronic      | 176504-176458  |
| 23 | rs10264092 | Intronic      | 59429-59379  |
| 24 | rs10276256 | Intronic      | 2801- 2751   |
| 25 | rs3757510  | Intronic      | 143990-144040  |
| 26 | rs9472     | Intronic      | 142915-142865  |

### **3.3 PRIMER DESIGN AND TESTING**

Polymerase chain reaction (PCR) is a commonly used method for amplification of short stretches of DNA. This method was used in this study to amplify the DNA around and including the SNP of interest prior to pyrosequencing.

PCR primers were designed using the Integrated DNA technologies Primer quest website (IDT), found at the following URL: <http://biotools.idtdna.com/Primerquest/>. This website was used to design both the forward and reverse PCR primers. In order to create amplified samples

suitable for pyrosequencing, the final PCR amplified product was restricted in size and was designed to be less than 500 base pairs. Once the ordered primers were received from IDT, they were diluted in H<sub>2</sub>O to a final concentration of 20pmol/ul. Of the 26 total SNPs identified, working PCR primers were designed for 24 SNPs.

In order to determine the best temperature for PCR primer annealing, gradient PCR was applied at temperatures ranging from 53 °C to 64 °C for each individual primer set. A total volume of 25ul was used for each PCR reaction. This included: 12.5ul of 2x Buffer I- obtained from *TaKaRa LA Taq*<sup>TM</sup> (which contains 2x 5mM Magnesium concentration), 2 ul of DNTPs also obtained from *LA Taq*<sup>TM</sup>, 0.5ul of forward primer 20pmol/ul, 0.5ul of reverse primer 20pmol/ul, 0.25ul of Taq polymerase, and 10ul of control human DNA concentration 50ng/ul. The PCR conditions were as follows: Pre-PCR conditions 50°C- 10 minutes, 96°C- 10 minutes, 32 PCR cycles: 96°C- 30 seconds (denaturation), gradient specific temperature- 30 seconds (annealing), 72°C- 45 seconds (extension), and two final cycles- 96° C for 10 min and 4° C for infinity

After PCR was completed, 13ul of each DNA sample was loaded into the individual wells of a 1.5% agarose gel. Gel electrophoresis was performed to assess the size and purity of the PCR product. It was also used to determine the best annealing temperature for each primer set. For each primer set the PCR product was then isolated from the agarose gel and was subject to HPLC sequencing. This step was utilized in order to assure the resulting product after PCR was indeed the expected product.

After PCR primers were designed using the IDT software, pyrosequencing primers were designed using the SOP<sup>3</sup> web based program designed by Massimo Trucco's group (107). The original version of the program, which was used in our current study, can be accessed via the URL: <http://biodev.hgen.pitt.edu/sop3/>. The program has since been updated and can now be accessed at the URL: <http://imgen.ccbb.pitt.edu/sop3/>. These primers were designed to allow sequencing of 10-20bp of DNA total around and including the SNP of interest. This program places a sequencing primer in the best area around the SNP region. The sequencing primer is placed at a minimum distance of 1 nucleotide proximally to the SNP, and at a maximum distance of 30 nucleotides from the SNP. This primer design software is designed to prevent the creation of a pyrosequencing primer that will form a primer dimer or hairpin loops in order to avoid the formation of a secondary structure that could interfere with primer annealing. This program

designs pyrosequencing primers located on both the 5' and 3' side of the SNP, and the one closest to the SNP of interest is selected by the program.

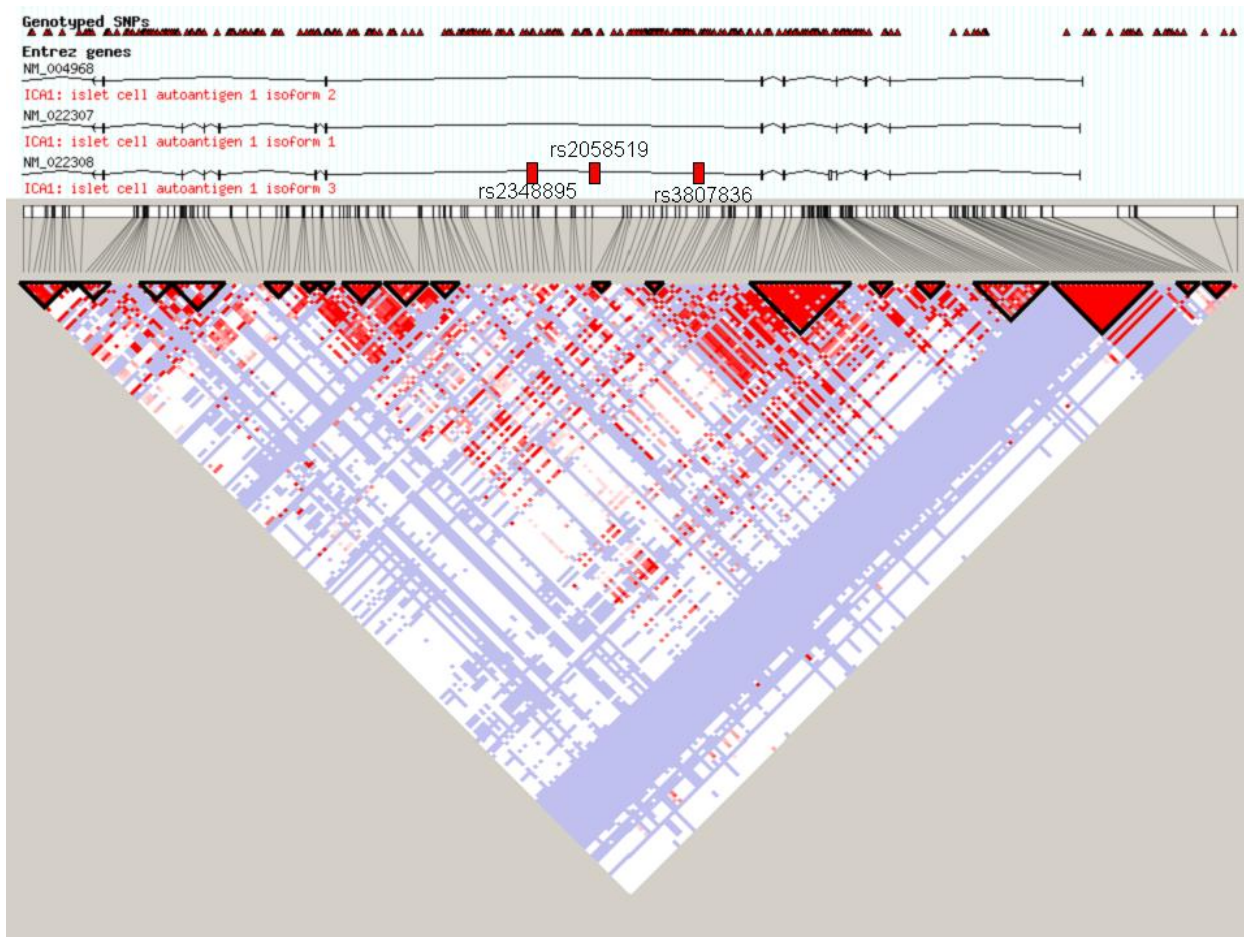
A single-stranded DNA template is necessary for pyrosequencing. This is essential because the pyrosequencing technology works by pairing each nucleotide to its complementary nucleotide on the DNA template strand. The method used to isolate the single-stranded DNA from double stranded DNA form was the use of a streptavidin bead. One of the two PCR primers was ordered with a biotin phosphoramidate attached as the 5' residue. Each streptavidin bead can then be used to isolate a single strand of DNA (details discussed in a section 3.6) to be used for the sequencing process. The PCR primer that was selected to be biotinylated was the PCR primer on the opposite side of the SNP from designed sequencing primer. For example if a sequencing primer was placed a few bases upstream of the SNP, the reverse PCR primer was biotinylated, and the forward primer was not; example (Figure 3).

ATGAGTCCTTAGAACTACAGCCATGTCTCCTCTGGGTAAAT**ATGAAAGGTTTCAG**  
**CCC****ACTCGG****A**GAGTAAGGACTTTCCAAATGTTCTTAATCTGACTGCGGAAAAAAAAA  
 AAAACTGGAAGAACTGAGAATTCAATTTACTTTCTTAGCCTTAAGGCTGTGTATAA  
 GCTACCCTGTTATTATCATTATTAATAATAAACACAGTGAATGTCAGACAGGCACTT  
 CATGGTAGATTTTTATATGAATTTTGTGCAAAAGTAAAGGGTTGCCAGGCTACTGCA  
 TAATATTACTCAGTTCTGGTCCAG**G**AGTACAAAATTCCACCAGGCCTCCTGAGCCCG  
 TCCCTCCAATTAGTGGCTTACTTCATTACGAATAAAGAACCCTCATTACTGTGGGCCT  
 TCAAGTGAAATAGGATCCCCTACCTCTTGAAATCCGCCTTTCTAAGTACAGACAGTT  
 CCCAACTTACAAACAATCCACTTA**AAGAACCACCAAGGCATT****CAGAGG**CTGCTGTAGG  
 CTCCTGCAGCTGCTGCCTGGGCCCCCTGCCAGCCCCCTGCCAGCCCCCTCCCCA  
 CACTCCTGGTCTGCTCAAGGGAGCGCTGCCTACCCCCCTTGCAAGGCCAGCTGCCCA  
 CCTGCCCTCGGACTGCAGGCTCGGAGCTGGGGGTT**CAGTGAGGCCTCAGA**

**Figure 3. Primer Design for SNP rs3807830: Forward and reverse primers in bold, biotinylated primer in bold italics, sequencing primer highlighted in grey, SNP nucleotide underlined and bold.**

### 3.4 SNP LOCATIONS

Sequencing primers were developed for a total of three SNPs. The location of these three SNPs is shown in Figure 3. All three SNPs are located within a central large intron in the *ICA1* gene.



**Figure 4.** Location of the three SNPs analyzed in the study. This picture shows the three isoforms of the *ICA1* gene; it is adapted from the HapMap2 database.

### 3.5 PCR AMPLIFICATION FOR SEQUENCING PRODUCTS

The final PCR annealing temperature that seemed to universally work best for all SNPs was 62 degrees C. Sequencing primers were created for the three SNPs illustrated in Figure 4. Further sequencing primer design for the additional 23 SNPs is currently in progress.

In order to prevent contamination between PCR runs, which could influence sequencing data, a Uracil N- Glycosylase enzyme was used. This enzyme was obtained from *GeneAmp*. It is known that uracil can be used as a replacement of thymine in the genetic code, as seen in the RNA product of a DNA template. This enzyme works by degrading template DNA containing

uracil bases, but it does not degrade DNA containing thymine bases. This enzyme exhibits optimal activity at 52 ° C and is irreversibly deactivated at 96° C. In this experiment PCR amplification of the template strand was achieved using dUTPs (also obtained from *GeneAmp*) with uracil nucleotides, rather than dNTPs with thymine nucleotides. Therefore all PCR products made from the template strand will contain uracil rather than thymine. Using this principle the original DNA template strand was combined with the enzyme at 52° C prior to the PCR amplification. This original template DNA strand, containing only thymine, was therefore not degraded by the enzyme. However, any possible PCR product contamination from previous PCR amplifications contained uracil bases only, and was therefore degraded prior to the amplification of the true template strand. Next, the PCR mixture was subject to a 96° C cycle prior to amplification to deactivate the glycosylase enzyme.

Based on this information, the ideal PCR amplification for these SNPs were applied. This consisted of two pre-PCR cycles-52° C for 10 min, 96° C for 10 min, 32 cycles of PCR- 96° C for 30 seconds (denaturation), 62° C for 30 seconds (annealing), 72° C for 45 seconds (extension), and two final cycles- 96° C for 10 min and 96° C for infinity. This last 96° C is applied to assure that if there is any remaining glycosylase enzyme it is denatured during this step, so as not to degrade our new PCR product. The sample were immediately removed after PCR amplification and frozen at -30° C. The PCR mixture consisted of a 25ul reaction containing 12.5ul *TaKaRa* Buffer I, 2ul *GeneAmp* DUTPs, 0.5ul of forward primer 20pmol/ul, 0.5ul of reverse primer 20pmol/ul, 10ul H<sub>2</sub>O, 0.25ul of Taq polymerase (Taq was purified by Dr. Steve Ringquist), and 0.25ul Uracil-N- Glycosylase enzyme. PCR reactions were performed in a 96 well plate format. This same 96 well plate format could also be used for pyrosequencing. See Table 5 for the designed PCR specific primer sequences.

**Table 5. PCR and Sequencing Primers for each SNP.**

| SNP        | Forward Primer (5' to 3')     | Reverse Primer (5' to 3')       | Sequencing Primer(5' to 3') |
|------------|-------------------------------|---------------------------------|-----------------------------|
| rs3807830  | ATGAAAGGTTTCAGCCCACTCGGA      | CCTCTGAATGCCTTGGTGGTTCTT-Biotin | GTTCTGCTGCA                 |
| rs2348895  | GAGGTTGCAGTGAGCCAAGATTG       | GCGCCAGGTTGCTAGGCAATATAA-Biotin | GAAGTTAAAATTACA             |
| rs2058519  | AACACAAACACGTCCACTCTTGGC      | CCAAGGCCCTGTTTAACCA- Biotin     | TTAAACCAAGGC                |
| rs3807852  | GCCTAACCATATCACAAGCCAAATACC   | AGTGGTGCTCTAAGGAAAGCAGGT        | N/A                         |
| rs3823838  | CACAAGAGTTGGTGGGCAAGACAA      | TTGGCAGCCTTACTCATTTCAGC         | N/A                         |
| rs3757524  | ACCATGTCCATGCCACAGCACTAT      | ACTCTCTTGTCTCACCCTACACT         | N/A                         |
| rs3757522  | GTTTGAAAGGAAGTGGTGGCTACTGA    | ACTGATGAGGATTGCGAGGCTTCA        | N/A                         |
| rs874721   | TCAGGTAATTTCCAGCAGGAAGGAGCA   | CATCTTGTGCTTTGGGTCTGTAGC        | N/A                         |
| rs3823833  | CCATAGGAGACCAAGAGCTGGAAA      | TATCCCTGAAGTTGCTGCCTCCAT        | N/A                         |
| rs3823832  | CGTTGTTCCAGGCATGAAATATCATCAAA | CTGTGGTGGCTGCAAAAC              | N/A                         |
| rs3807824  | GATGGATCCAGGATTACCAGCTCA      | ATGAAGGGCACGGACATAAGAGCA        | N/A                         |
| rs3757515  | AGGTTGAGGTTTCACCATACGCCT      | TTGGTGGATTTCCTCAGGCACTA         | N/A                         |
| rs3807816  | TGATTCTACAACCACACAGCACTC      | CAAAGATCTGCTTGCTTGGTGGGT        | N/A                         |
| rs1008038  | TAATGCCTGGCTTTGGACCATGAG      | AGAAGCAGAGGCCAAGGATGAAT         | N/A                         |
| rs6463765  | TGAGGCCAAGGGCTCCATAAATCT      | TATATCTGGCCCTGAAGTGTGGCA        | N/A                         |
| rs887848   | ACTGCACAAGGAACGAGAGA          | AGGAGGACCTGGAGAAAGTCAGAGAA      | N/A                         |
| rs2110332  | TTTCCCATCTTGCTCCCTACCATC      | GGCATGCTTAGTGAAGCTTTGCAG        | N/A                         |
| rs1981715  | TTTCTGGTGGGAGAAACAGAGCCT      | TTCTGTGGCCCTTCCTTGGACAAT        | N/A                         |
| rs3807818  | TGCTTCTTCATCTCCCTACCCAA       | TAGGTAGGACAGTGTGAGCAGGTT        | N/A                         |
| rs759262   | AGGTGTGGTGGCAAGTGCCTATAA      | CCCACCTCCAATGTCAGCCAAA          | N/A                         |
| rs3807849  | TGTCCAGAGCAAGAATGCCACTCA      | CGTGAGGTTTGTGCCATTGCTTA         | N/A                         |
| rs9472     | GGGAGTGAGTTTCTCTCTCTTAC       | TAACTGGAAGCAAGCAGTAGGCA         | N/A                         |
| rs10276256 | GCTCTTTCTTCTTATCTCGGCAAGC     | CCACAACCCACACACATTCGAT          | N/A                         |
| rs3757510  | CAGCGACGACTGAATGAAGTT         | GCCATTTGTAGCAGTTTCATCTTCTC      | N/A                         |

### 3.6 DNA SEQUENCING AND GENOTYPE DETERMINATION

Genotypes were determined using the four-enzyme pyrosequencing technology which is incorporated in the commercial PSQ<sup>TM</sup> HS 96 system. The PSQ system incorporates the enzymes DNA polymerase, ATP sulfurylase, luciferase, and apyrase. The pyrosequencing process begins with the isolation of the biotin bead along with the attached single stranded DNA template. For the isolation of each individual PCR product 5ul of PCR reaction was combined in the following solution: 38ul binding buffer (containing 10mM Tris-HCL, 2M NaCl, 1mM EDTA, 0.1% Tween<sup>TM</sup> 20), 35ul H<sub>2</sub>O, and 2ul of Streptavidin Sepharose<sup>TM</sup> High Performance beads *Amersham Biosciences*. The solution was mixed on a shaker for 15 minutes at 23° C and 14000rpm. Immediately after this mix, the single stranded DNA attached to the streptavidin beads was isolated using a vacuum prep tool. The bead/DNA combination was then washed for 30 seconds in 70% ethanol, it was denatured in a 0.2M NaOH solution, and it was washed in a wash buffer (containing 10mM Tris, pH 7.6). Next, following the wash, the single-stranded DNA

was combined with the sequencing primer solution containing 10.8ul 10x annealing buffer (containing 20mM Tris-Acetate and 2mM Mg-Acetate), and 1.2ul of the pyrosequencing primer (concentration 3pmol/ul). The final solution was then placed on a hot plate for 2 minutes to allow the sequencing primer to hybridize to the single stranded DNA template. The solution was allowed to cool for 15 minutes prior to the sequencing step.

This final DNA solution was placed in 96 well plate format into the pyrosequencing machine purchased from the *Biotage Corporation*. The reagents were then added into the machine. The reagents that were used included the four enzymes and the nucleotides, all of which were included in the Pyro Gold CDT reagents made by *Biotage*. Repeated cycles of nucleotide addition were performed within the pyrosequencing machine which allowed each nucleotide to incorporate into the DNA strand only if it was complementary to the base in the template strand. Synthesis of the DNA was accompanied by the release of a pyrophosphate equal in molarity to that of the incorporated nucleotide. The pyrophosphate (PPi) was converted to ATP by the enzyme ATP- sulfurylase. The ATP then reacted with the luciferase enzyme and it was converted into light. This light was recorded in real-time signals, and the amount of light produced was proportional to the amount of ATP generated and was therefore proportional to the number of nucleotides incorporated. The light was focused into a highly sensitive CCD (charge coupled device) camera within the pyrosequencing machine. After each nucleotide addition, non-incorporated nucleotides were degraded by the apyrase enzyme. The cycle was then repeated for the addition of each individual nucleotide. Only one specific nucleotide was released at a time. The order in which the nucleotides were dispensed was based on the specific DNA sequence to be analyzed, and therefore varied from SNP to SNP. For the specific dispensation order refer to Table 6.



**Table 6. Pyrosequencing Assay Information.**

| SNP       | Sequence to be analyzed | Dispensation order |
|-----------|-------------------------|--------------------|
| rs3807830 | GC/GAGTACAAAATTCCA      | TCGCAGTAC          |
| rs2348895 | CCC/TTCCTTCAATCT        | GCTGCTCAT          |
| rs2058519 | CAGA/GGAGAAATACACGTT    | GCAGAGAGAGT        |

The real time data output from the CCD camera was then visualized graphically in the form of a pyrogram, as the data was analyzed automatically by the computer. The data could have also been manually analyzed by a user on the computer (110). The peak heights seen in the pyrogram were directly proportional to the number of nucleotides incorporated. If a nucleotide was dispensed that was not complimentary to the next nucleotide in the template strand, no light was produced, and no peak was seen on the pyrogram. The nucleotide sequence could then be visualized by the peaks in the pyrogram. A new peak was created as each nucleotide was bound to its complementary base on the DNA strand. If more than one of the same nucleotide was added, for example two adenines in a row, the peak produced was 2 times the height of a single peak, rather than 4 separate peaks. The principal of the pyrosequencing process is shown Figure 4 (110).

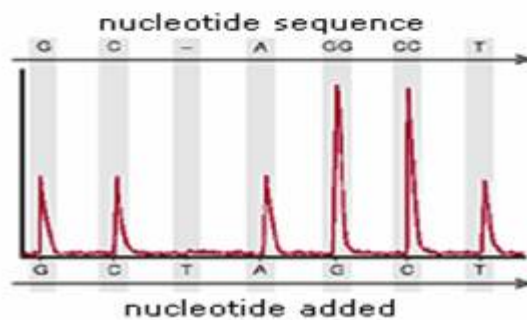
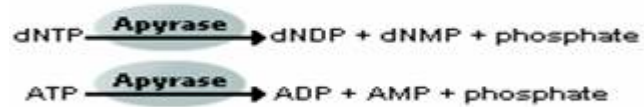
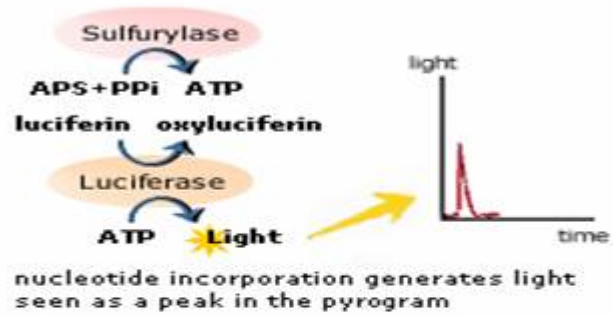


Figure 5. The Principle of Pyrosequencing (110).

The PSQ<sup>TM</sup> HS 96 system automatically analyzed our data. The computer represents its data analysis using three color calls. A blue call indicated that the pyrosequencing run passed; this means that the area around the SNP (“sequence to be analyzed”) had the expected nucleotides in the expected order. It also means that there was a low signal-to noise ratio, meaning that there was not excess background light present besides the true signal. A yellow call indicated that the run is questionable, which means that the “sequence to be analyzed” was not exactly as expected by the data input in the computer system. In this case the SNP genotype data may have still been correct, but there may have been some difficulty in sequencing in the area around the SNP. *Biotage* suggests that yellow calls should be manually analyzed before the genotype data is used. The last type of call is a red call, which indicated a failed run. This meant that the data obtained by pyrosequencing was much different expected. This may have been due to an error in the PCR process, an error in primer design, or a high background to noise ratio. In our experiment all SNP analysis were performed in duplicate. Only samples that had two blue calls made by the pyrosequencing machine were used for our results. If there was an error in sequencing producing a yellow or red call, the PCR and pyrosequencing steps were repeated to attempt to obtain two passing sequencing runs. If two passing runs (blue) were not obtained the sample data was not used, and is recorded in the result section as “no results”.

### **3.7 STATISTICAL METHODS**

The allele frequencies were calculated by manual allele counting. The chi-squared test for Hardy-Weinberg equilibrium (HWE) was applied to assess the deviations of the observed genotypes from the expected HWE allele frequencies.

Fisher’s exact test was used to determine whether there was a difference in the observed vs expected allele frequencies between the control population and the diabetic cases. This test was applied using 2x2 tables to analyze the difference between the common homozygote vs the heterozygote plus the rare homozygote in both the controls and the pooled diabetic cases. The

test was also performed to compare the common homozygote plus the heterozygote vs the rare homozygote for SNP rs2058519.

Fisher's exact test was used to analyze the data using 3x2 tables to assess the difference between the three possible allele types. This test was again performed comparing the 106 pooled diabetic cases to the non-diabetic controls.

Fisher's exact test was also applied to analyze the original 73 diabetic cases versus the extra 33 diabetic cases. This test was performed to establish that there is no significant difference between these two populations to support pooling the two sample populations.

The HWE analysis was performed using the web-based calculator located at the URL: <http://www.genes.org.uk/software/hardy-weinberg.shtml>. These results were confirmed by performing the calculations by hand. All other statistical analysis was performed using the statistical program STATA 7.0.

## 4.0 RESULTS

### 4.1 GENOTYPE COMPARISON AT THREE SNP LOCI

The genotype and allele frequencies established for the three *ICAI* SNPs are shown in Tables 7-9. Although we were not able to obtain genotype data for 100% of the individuals we tested, we were able to sequence 99.96% of our total samples.

**Table 7. Genotype data for rs3807830 in diabetics and controls.**

| <b>Genotype</b> | <b>Controls</b> |      | <b>Original Diabetics</b> |      | <b>Pooled Diabetics</b> |      |
|-----------------|-----------------|------|---------------------------|------|-------------------------|------|
|                 | n               | %    | n                         | %    | n                       | %    |
| G/G             | 69              | 87.3 | 58                        | 79.5 | 88                      | 83   |
| C/G             | 10              | 12.7 | 9                         | 12.3 | 12                      | 11.3 |
| C/C             | 0               | 0    | 0                         | 0    | 0                       | 0    |
| No results      | 0               | 0    | 6                         | 8.2  | 6                       | 5.7  |
| Total           | 79              |      | 73                        |      | 106                     |      |

**Table 8. Genotype data for rs2348895 in diabetics and controls.**

| <b>Genotype</b> | <b>Controls</b> |      |  | <b>Original Diabetics</b> |      | <b>Pooled Diabetics</b> |      |
|-----------------|-----------------|------|--|---------------------------|------|-------------------------|------|
|                 | n               | %    |  | n                         | %    | n                       | %    |
| C/C             | 68              | 86.1 |  | 60                        | 82.2 | 88                      | 83   |
| C/T             | 9               | 11.4 |  | 8                         | 11   | 11                      | 10.4 |
| T/T             | 1               | 1.2  |  | 0                         | 0    | 0                       | 0    |
| No results      | 1               | 1.2  |  | 5                         | 6.8  | 7                       | 6.6  |
| Total           | 79              |      |  | 73                        |      | 106                     |      |

**Table 9. Genotype data for rs2058519 in diabetics and controls.**

| <b>Genotype</b> | <b>Controls</b> |      | <b>Original Diabetics</b> |      | <b>Pooled Diabetics</b> |      |
|-----------------|-----------------|------|---------------------------|------|-------------------------|------|
|                 | n               | %    | n                         | %    | n                       | %    |
| A/A             | 36              | 44.3 | 42                        | 57.5 | 64                      | 60.4 |
| A/G             | 36              | 45.6 | 24                        | 32.9 | 32                      | 30.2 |
| G/G             | 7               | 8.9  | 1                         | 1.4  | 1                       | 0.94 |
| No results      | 1               | 1.2  | 6                         | 8.2  | 9                       | 8.4  |
| Total           | 79              |      | 73                        |      | 106                     |      |

The comparisons of the original diabetic samples to the additional diabetic samples are shown in Tables 10-12. The large p-values for these comparisons indicate that there is no evidence to suggest a statistical difference between the original and additional diabetic samples.

**Table 10. Comparison of original diabetic samples versus additional diabetic samples for SNP rs3807830.**

|                    | G/G | C/G | C/C | Total |
|--------------------|-----|-----|-----|-------|
| Original Diabetics | 58  | 9   | 0   | 67    |
| Extra Diabetics    | 30  | 3   | 0   | 33    |
| Total              | 88  | 12  | 0   | 100   |

Fisher's exact= 0.746

**Table 11. Comparison of original diabetic samples versus additional diabetic samples for SNP rs2348895.**

|                    | C/C | C/T | T/T | Total |
|--------------------|-----|-----|-----|-------|
| Original Diabetics | 60  | 8   | 0   | 68    |
| Extra Diabetics    | 28  | 3   | 0   | 31    |
| Total              | 88  | 11  | 0   | 99    |

Fisher's exact= 1.000

**Table 12. Comparison of original diabetic samples versus additional diabetic samples for SNP rs2058519.**

|                    | A/A | A/G | G/G | Total |
|--------------------|-----|-----|-----|-------|
| Original Diabetics | 42  | 24  | 1   | 67    |
| Extra Diabetics    | 22  | 8   | 0   | 30    |
| Total              | 64  | 32  | 1   | 97    |

Fisher's exact= 0.644

The genotype frequencies of each *ICAI* SNP were compared between the pooled diabetic cases and control subjects using 3x2 tables (Tables 13-15). The p-value of 1.00 (Fisher's exact test) obtained for SNP rs3807830, comparing the pooled cases vs the controls, shows *no significant difference* between the cases and controls (Table 13).

The p-value of 0.707 (Fisher's exact test) for rs2348895 comparing the pooled cases vs the controls shows *no significant difference* between the cases and controls (Table 14).

There appears to be a *significant difference* between the diabetic cases and the controls for SNP rs2058519 (Table 15). A p-value of 0.003 (Fisher's exact test) was obtained when comparing the control samples to the pooled diabetic samples at this SNP. This value is less than the cutoff value of 0.05 that we are using in this study to determine statistical significance and therefore there is a statistically significant difference between the two groups at this SNP.

**Table 13. 3x2 Table comparison of pooled cases vs control samples for SNP rs3807830.**

|              | G/G | C/G | C/C | Total |
|--------------|-----|-----|-----|-------|
| Pooled Cases | 88  | 12  | 0   | 100   |
| Controls     | 69  | 10  | 0   | 79    |
| Total        | 157 | 88  | 12  | 100   |

Fisher's exact =1.000

**Table 14. 3x2 Table comparison of pooled diabetic cases vs control samples for SNP rs2348895.**

|              | G/G | C/G | C/C | Total |
|--------------|-----|-----|-----|-------|
| Pooled Cases | 88  | 11  | 0   | 99    |
| Controls     | 68  | 9   | 1   | 78    |
| Total        | 156 | 20  | 1   | 177   |

Fisher's exact= 0.707

**Table 15. 3x2 Table comparison of pooled diabetic cases vs control samples for SNP rs2058519.**

|              | A/A | A/G | G/G | Total |
|--------------|-----|-----|-----|-------|
| Pooled Cases | 64  | 32  | 1   | 97    |
| Controls     | 35  | 36  | 7   | 78    |
| Total        | 99  | 68  | 8   | 175   |

Fisher's exact= 0.003

The genotype frequencies of each of the *ICAI* SNPs were also compared between the pooled diabetic cases and control subjects using 2x2 tables. For this analysis, the individuals who had the common homozygous genotype were compared to individuals who had either the heterozygous genotype or the rare homozygous genotype (Tables 16-18). The comparison was also assessed using Fisher's exact test. For SNP rs3807830 there appeared to be no difference between the cases and the controls  $p= 1.00$ . For SNP rs2348895 there also appeared to be no difference between the diabetic cases and the controls samples  $p= 0.816$

There again appeared to be a *significant* difference between the diabetic cases and the control samples at SNP rs2058519. Comparing the pooled diabetic cases to the non-diabetic controls there was a p-value of 0.004 (Table 18). The odds ratio for this comparison was 2.38 and the 95% CI was 1.2 to 4.6. This indicates that individuals with the AA genotype are 2.38 times more likely to be affected with type 1 diabetes compared to those with the AG or GG genotype.

A comparison of genotype frequency was also performed for SNP rs2058519 to compare individuals with the common homozygote or the heterozygous genotype to individuals with the



rare homozygote (Table 19). There again appears to be a significant difference between cases and controls with this assessment at this SNP location  $p=0.023$ .

**Table 16. 2x2 Table comparing individuals with the common homozygote vs the rare homozygote or heterozygous individuals for SNP rs3807830.**

|                              | GG  | CG/CC | Total |
|------------------------------|-----|-------|-------|
| Pooled Cases                 | 88  | 12    | 100   |
| Controls                     | 69  | 10    | 79    |
| Total                        | 157 | 22    | 179   |
| Fisher's exact p-value= 1.00 |     |       |       |

**Table 17. 2x2 Table comparing individuals with the common homozygote vs the rare homozygote or heterozygous individuals for SNP rs2058519.**

|                              | CC  | CT/CC | Total |
|------------------------------|-----|-------|-------|
| Pooled Cases                 | 88  | 11    | 99    |
| Controls                     | 69  | 10    | 79    |
| Total                        | 157 | 21    | 178   |
| Fisher's exact p-value= 1.00 |     |       |       |

**Table 18. 2x2 Table comparing individuals with the common homozygote or heterozygous individuals vs those with the rare homozygote for SNP rs2058519.**

|  | AA | AG/GG | Total |
|--|----|-------|-------|
| Pooled Cases   | 64 | 33    | 97    |
| Controls   | 35 | 43    | 78    |
| Total  | 99 | 76    | 175   |
| Fisher's exact p-value= 0.004<br>Odds Ratio= 2.383<br>95% CI= 1.23-4.6 |    |       |       |

**Table 19. 2x2 Table comparing individuals with the common homozygote or heterozygous individuals vs those with the rare homozygote for SNP rs2058519.**

|                               | AA/AG | GG | Total |
|-------------------------------|-------|----|-------|
| Pooled Cases                  | 96    | 1  | 97    |
| Controls                      | 71    | 7  | 78    |
| Total                         | 167   | 8  | 175   |
| Fisher's exact p-value= 0.023 |       |    |       |

## 4.2 HWE ANALYSIS OF GENOTYPING DATA

All samples appeared to be in Hardy-Weinberg Equilibrium (HWE) using the chi-squared test for HWE (Table 20). None of the chi-squared p-values obtained from this test was significant. This means that the null hypothesis that the samples are in HWE cannot be rejected. We do not have enough evidence to suggest that the samples are not in HWE. Based on this preliminary information we would expect that additional samples would also be in HWE.

**Table 20. HWE chi-squared p-values for the three SNPs studied.**

|                         | rs3807830 | rs2348895 | rs2058519 |
|-------------------------|-----------|-----------|-----------|
| Controls                | P=0.36    | p=1.12    | p=0.28    |
| Pulled Diabetic Cases   | P=0.41    | p=0.34    | p=1.93    |
| Original Diabetic Cases | p=0.35    | p=0.27    | p=1.41    |

Ho=Samples are in HWE

## **5.0 DISCUSSION**

### **5.1 SNP SIGNIFICANCE**

In the gene *ICA1* at SNP rs2058519 there appears to be a clear difference in the genotype data between the control individuals and the diabetic cases. This is an intronic SNP and thus located in an untranslated region of the gene. Our data indicates that the odds that individuals with the AA genotype have type 1 diabetes are 2.383 times more likely than individuals with the AG or GG genotype. However, the data clearly show that not all individuals with the AA genotype have type 1 diabetes, and therefore as expected, this is not a dominant locus. This finding appears to support our hypothesis that SNPs within the *ICA1* gene contribute to the onset to type 1 diabetes. The other two SNPs analyzed in this study rs3807830, and rs2348895 showed no significant difference, and the genotype at these locations does not appear to significantly differ between diabetic and non-diabetic individuals.

### **5.2 POSSIBLE FUNCTIONAL RELATIONSHIP OF AN INTRONIC POLYMORPHISM**

Although this *ICA1* SNP is located in an intronic region, it is possible that it still plays a functional role and may affect alternative splicing or gene transcription. Further studies are necessary to determine the functional role that SNP rs2058519 may play in the disease process. There is currently no published data suggesting that SNPs within the non-coding regions of the *ICA1* gene in humans or animal models are associated with type 1 diabetes. However, the

previously discussed unpublished data by Chang and Pietropaolo *et al.* suggests that a single SNP in non-coding promoter region of *ICA1* may significantly affect ICA69 transcription in the thymus. The decreased thymic expression could in turn contribute to negative *ICA1* selection in the thymus resulting in failure to destroy autoreactive T-cells directed against ICA69, and therefore may play a role in autoimmune diabetes pathogenesis.

A similar study was performed by Bassuny *et al.* studying the *FOXP3/Scurfin* gene, which is also believed to be associated with type 1 diabetes (111). Bassuny's group studied single nucleotide polymorphisms as well as microsatellite changes in the intronic and promoter regions of the gene. The study demonstrated that two microsatellite polymorphisms in two different introns were in significant linkage disequilibrium, and that there was a clear difference at these sites in diabetic versus control patients. It was also found that the genetic changes caused a significant difference in enhancer activity, and therefore the gene appears to play a role in the onset of type 1 diabetes in the Japanese population they were studying. The group's data supports our study design to scan the intronic regions of the *ICA1* gene to determine if there is a significant difference between cases and controls.

### **5.3 IDEAL SAMPLE SIZE**

The initial power calculations performed by a collaborating biostatistics group at CMU suggested that 400 non-diabetic control individuals and 400 diabetic cases would be necessary to provide >90% power to identify major haplotypes. Based on this information and the number of individuals we were able to genotype in this study, at least 322 additional control individuals and 303 additional type 1 diabetic individuals should be genotyped to provide data that reaches the appropriate power.

This study is a preliminary study to determine whether there was a difference between diabetics and controls at these three genetic loci. This research suggests a possible association between SNP rs2058519 and type 1 diabetes. Once the experiment has been conducted to its full power (>400 samples for both cases and controls) replication of the results will be important to assure us that the discovered associations are real. Once full power has been achieved, the Bonferroni correction should be applied. This technique should be applied to assess the

probability that a type 1 error is made when assessing the data. A type 1 error is the chance of incorrectly declaring a difference, or relationship as true simply due to chance. This test will support that the associations we find are true. Due to the fact that our study has currently not achieved its optimal N value, we feel that this correction measure is not applicable at the present time, but we support the use of this statistical method once the appropriate power has been achieved.

## **5.4 FUTURE STUDIES/APPROACH**

In the future it will be important to determine if there is a difference at the additional 23 SNP loci that have not yet been assessed in this preliminary experiment. Original tag SNPs were determined using the International HapMap 1 project (112). The HapMap is an international collaboration in which genetic data is gathered from African, Asian, and European countries. The goal of the group is to compare genetic sequence variants to aid in finding the genes involved in disease and drug therapy responses. Genetic variants will be compared to determine if there are any major genetic differences in various ethnic populations around the world. The 26 SNPs identified in this study were obtained using the first HapMap build. At the beginning of our study a total of 44 Tag SNPs were identified in the *ICAI* gene. In the most recently updated HapMap 2 build a total of 260 *ICAI* tag SNPs exist. Prior to the continuation of our project, the new build should be accessed to consider other possible SNPs of interest. Our group plans to use this build to identify any new “true” SNPs. All “true” SNPs identified in the *ICAI* locus will then be genotyped in both diabetic cases and control subjects.

Based on the fact that we identified one SNP in *ICAI* which may be associated with type 1 diabetes, we should consider continuing our screening by first focusing on SNPs that are located close to SNP rs2058519 to determine if there are several SNPs in linkage disequilibrium. It is also essential to incorporate information between multiple SNP loci to compare the haplotype frequency profiles between the cases and controls. The haplotype frequency will determine if several SNPs in the *ICAI* gene are inherited together. The overall goal of the project

is to determine whether there are undiscovered variants in the *ICAI* gene which are associated with type 1 diabetes susceptibility, and to determine if there is a difference in the haplotypes between the cases and controls.

In addition to the case control study, family-based association tests are also important to determine if associations can be identified between *ICAI* SNPs and type 1 diabetes. The most widely used family based association test is the Transmission Disequilibrium Test (TDT test) (113). This test first developed by Spielman *et al.* applies analysis using a proband and their two parents (114), occasionally this technique may also include unaffected siblings. (110). Although other techniques have been suggested to replace this technique (115), thus far it has been demonstrated that the original TDT test provides the greatest power (114). The main purpose of the TDT test is to assess the different transmission pattern of alleles from heterozygous parents to their diseased children. The goal of this test is to determine if there is a variation from the expected transmission frequency of 0.5 (116). The TDT test is capable of showing a valid association even if there is population stratification (114, 116). The test has been used to identify associations between several HLA-DQ genetic variants and type 1 diabetes using case-parent trios in various ethnic populations (113,116, 117, 118).

Complex traits, such as type 1 diabetes, are suspected to involve both multiple genetic as well as environmental factors. We would like to examine these gene-gene interactions by using logistic regression to determine if there is a significant difference between cases and controls when considering multiple loci. In particular, we are interested in determining whether the various HLA-DQ alleles interact with allelic differences in the *ICAI* gene.

Overall, we will need to analyze our data sets to determine if the current and future associations we find, such as the association we found in rs2058519, play a ‘true’ causal role in disease pathogenesis and if particular SNPs are in linkage disequilibrium with other variants.

## **5.5 GENOTYPING TROUBLE SHOOTING**

One major setback with the pyrosequencing technique is that genotype data are not always obtained for every sample in each run, which often requires pyrosequencing repetition in order to obtain valid data. Therefore, the technique is less time efficient than is theoretically expected.

There are several reasons a run could fail to produce genotype data, including ineffective PCR product, degradation of the PCR or pyrosequencing primers, and poor binding of the Streptavidin Sepharose to the PCR product (this could be due to improper suction with the tool used for this procedure). Another troubleshooting idea to consider is that we know the peak compensation level as well as the stringency level used by the pyrosequencing computer to determine if a run successfully passed can be altered to get a greater number of passing calls. However, we used the default factory settings with our data to assure that we would be using the most stringent testing and obtaining the most accurate data. It may be possible to alter these settings in order to obtain a greater percentage of passing calls with each run without jeopardizing the quality of the data.

## **5.6 OTHER POSSIBLE TECHNIQUES**

The genotyping method used in the study has proven to be an effective genotyping technique. However, other assays have been developed to analyze high-throughput methods for genotyping SNPs. One such method that should be considered is the TaqMan or 5' nuclease allelic discrimination assay (112). This assay has an error rate of less than 1 in 2000 genotypes. It uses one fluorescent reader and five 96 well PCR machines, and does not require automated pipetting. It is reported that a single individual can genotype over 1000 samples in one day, and that it is possible to test a new SNP association in less than 1 week (119).

## **5.7 SUMMARY**

This study has genotyped 3 SNPs within the promoter region of the *ICAI* gene. Differences between diabetes and control subjects were analyzed. A significant difference was found between the diabetic cases and control subjects at SNP rs2058519 (Fisher's exact p-value= 0.0003). No significant associations were found at the two other SNP locations rs3807830, or rs2348895. This project has provided additional information about the genetics of type 1 diabetes

and serves as the basis for further studies. Understanding the genetics associated with type 1 diabetes has an important public health significance because it will allow for improved prediction and treatment and may possibly aid in the development of a cure.



## BIBLIOGRAPHY

- 1.) Bell G, Polonsky K. (2001) Diabetes mellitus and genetically programmed defects in  $\beta$ -cell function. *Nature* 414:788-791.
- 2.) Pietropaolo M, LeRoith D. (2001) Pathogenesis of Diabetes: Our Current Understanding. *Clin Cornerstone* 4(2):1-16.
- 3.) Field L. (2002) Genetic linkage and association studies of type 1 diabetes: challenges and rewards. *Diabetologia* 45:21-35.
- 4.) Atkinson M, Maclaren N. (1994) The pathogenesis of insulin-dependent diabetes mellitus. *The New England Journal of Medicine*. 331:1428-1436.
- 5.) Mathis D, Vence L, Benoist C. (2001)  $\beta$ -Cell death during progression to diabetes. *Nature*. 414:792-798.
- 6.) Pietropaolo M, Becker D. (2001) Type 1 diabetes interventions trials. *Diabetes*. 2: 2-11.
- 7.) Miyazaki I, Cheung R, Gaedigk R, Hui M, Meulen J, Rajotte R, Dosch H. (1995) T cell activation and anergy to islet cell antigen in type 1 diabetes. *Journal of Immunology*. 154:1461-1469.
- 8.) Vaarala O, Knip M, Paronen J, Hamalainen A, Muona P, Vaatainen M, Ilonen J, Simell O, Hans A. (1999) *Diabetes*. 48:1389-1394.
- 9.) Pietropaolo M, Trucco M. (2003) Genetics of type 1 diabetes. Contemporary Endocrinology. Humana Press Inc. NJ. 23-54.
- 10.) Allen C, Palta M, D'Alessio D. (1991) Risk of diabetes in siblings and other relatives of IDDM subjects. *Diabetes* 40: 831-836.
- 11.) Dahlquist G, Blom L, Tuvemo T, Nystrom L, Sandstrom A, Walls S. (1989) The Swedish childhood diabetes study-results from a nine years case register and a one year case referent study indicating that type 1 (insulin-dependent) diabetes mellitus is associated with both type 2 (non-insulin-dependent) diabetes mellitus and autoimmune disorders. *Diabetologia* 32:2-6.

- 12.) Dahlquist G, Blom L, Lonnberg G. (1991) The Swedish childhood diabetes study- a multivariate analysis of risk determinates for diabetes in different age groups. *Diabetologia* 34:757-762.
- 13.) Tuomilehto J, Lounamaa R, Tuomilehto-Wolf E, Reunanen A, Virtala E, Kaprio E., Akerblom H, and the Childhood Diabetes in Finland (DiMe) Study Group. (1992) Epidemiology of childhood diabetes mellitus in Finland- background of a nationwide study of type 1 (insulin-dependent diabetes mellitus. *Dibetologia*. 25:70-76.
- 14.) Tillil H, Kobberling J. (1987) Age-correlated empirical genetic risk estimates for first degree relatives of IDDM patients. *Diabetes*. 36: 93-99.
- 15.) Mein C, Esposito L, Dunn M, Johnson G, Timms A, Goy J, Smith A, Sebag-Montefiore L, Merriman M, Wilson A, Pritchard L, Cucca F, Barnett A, Bain S, Todd J. (1998) *Nature Genetics*. 19:297-300.
- 16.) Steck A, Bugawan T, Maria A, Emery L, Blair A, Norris J, Redondo M, Babu S, Erlich H, Eisenbarth G, Rewers M. Association of Non-HLA Genes With Type 1 Diabetes Autoimmunity (2005) *Diabetes*. 54:2482-2486.
- 17.) Lernmark Å, Ott J. (1998). Sometimes it's hot, sometimes it's not. *Nature Genetics*. 19: 213-214.
- 18.) Pietropaolo M, Becker D, LaPorte R, Dorman J, Riboni S, Rudert W, Mazumdar S, Trucco M. (2002) Progression to insulin-requiring diabetes in seronegative prediabetic subjects: the role of two HLA-DQ high-risk haplotypes. *Diabetologia*. 45:66-76.
- 19.) Warram J, Krolewski, A, and Kahn, C. (1988) Determinants of IDDM and perinatal mortality in children of diabetic mothers. *Diabetes*. 37:1328-1334.
- 20.) Nepom G. (1995) Class II antigens and disease susceptibility. *In Annu Rev Med*. 17-25.
- 21.) Warram J, Krolewski A, Gottlieb M, Kahn C (1984) Differences in risk of insulin-dependent diabetes in offspring of diabetic mothers and diabetic fathers. *N.Engl.J.Med*. 311:149-152.
- 22.) Dorman J, McCarthy B, O'Leary A, Koehler A. (1995) Risk factors for insulin-dependent diabetes. In Diabetes in America. National Institute of Diabetes and Digestive and Kidney Diseases. 165-177.
- 23.) Vadheim C, Rotter J, Maclaren N, Riley W, Anderson C. (1986) Preferential transmission of diabetic alleles within the HLA gene complex. *N.Engl.J.Med*. 315:1314-1318.
- 24.) Björnses P, Aaltonen J, Vikman A, Perheentupa J, Ben-Zion G, Chiumello G, Dahl N , Heideman P, Hoorweg-Nijman J, Mathivon L, Mullis P, Pohl M, Ritzen M, Romeo G, Shapiro M, Smith C, Solyom J, Zlotogora J, Peltonen L. (1996) Genetic homogeneity of autoimmune polyglandular disease type I. *Am J Hum Genet*. 59:879-886.

- 25.) Aaltonen J, Horelli-Kuitunen N, Fan J, Perheentupa J, Myers R, Palotie A, Peltonen L. (1997) High-resolution physical and transcriptional mapping of the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy locus on chromosome 21q22.3. *Genome Res.* 7:820-829.
- 26.) Bennett C, Yoshioka R, Kiyosawa H, Barker D, Fain P, Shigeoka A, Chance P. (2000) X-Linked syndrome of polyendocrinopathy, immune dysfunction, and diarrhea maps to Xp11.23-Xq13.3. *Am J Hum Genet.* 66:461-468.
- 27.) Redondo M, Eisenbarth G. (2002) Genetic control of autoimmunity in Type I diabetes and associated disorders. *Diabetologia.* 45:605-622.
- 28.) Pitkänen J, Doucas V, Sternsdorf T, Nakajima T, Aratani S, Jensen K, Will H, Vähämurto P, Ollia J, Vihinen M, Antonarakis S, Kudoh J, Shimizu N, Krohn K, Peterson P. (2000) The autoimmune regulator protein has transcriptional transactivating properties and interacts with the common coactivator CREB-binding protein. *J Biol Chem.* 275 :16802-16809.
- 29.) Wen L, Wong F, Tang J, Chen N, Altieri M, David C, Flavell R, Sherwin R. (2000) In vivo evidence for the contribution of human histocompatibility leukocyte antigen (HLA)-DQ molecules to the development of diabetes. *J Exp Med.* 191:97-104.
- 30.) Miyazaki T, Uno M, Uehira M, Kikutani H, Kishimoto T, Kmoto M, Nishimoto H, Miyazaki J. (1990) Direct evidence for the contribution of unique I-A<sup>NOD</sup> to the development of insulinitis in non-obese diabetic mice. *Nature.* 345:722-724.
- 31.) Slattery R, Kjer-Nielsen L, Allison J, Charlton B, Mandel T, Miller J. (1990) Prevention of diabetes in non-obese diabetic I-Ak transgenic mice. *Nature.* 345:724-726.
- 32.) Lund T, O'Reilly L, Hutchings P, Kanagawa O, Simpson E, Gravely R, Chandler P, Dyson J, Picard J, Edwards A, Kioussis D, Cook A. (1990) Prevention of insulin-dependent diabetes mellitus in non-obese diabetic mice by transgenes encoding modified I-A beta chain or normal I-E alpha chain. *Nature.* 345:727-729.
- 33.) Boitard C, Bendelac A, Richard M, Carnaud C, Bach J. (1988) Prevention of diabetes in nonobese diabetic mice by anti-I-A monoclonal antibodies: transfer of protection by splenic T cells. *Proc.Natl.Acad.Sci.(USA)* 85:9719-9723.
- 34.) Brown J, Jardetzky, Gorga J, Stern L, Urban R, Strominger J, Wiley D. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature.* 364:33-39.
- 35.) Lee K, Wucherpfennig K, Wiley D. (2001) Structure of a human insulin peptide-HLA- DQ8 complex and susceptibility to type 1 diabetes. *Nat Immunol.* 2:501-507.
- 36.) She J, Marron M. (1998) Genetic susceptibility factors in type 1 diabetes: linkage, disequilibrium and functional analyses. *Curr Opin Immunol.* 114:370-376.

- 37.) Pietropaolo M, Trucco M. (2000) Major histocompatibility locus and other genes that determine the risk of development of Type 1 diabetes mellitus. *Diabetes Mellitus. A fundamental and clinical text*. D.LeRoith, Taylor,S.I., and Olefsky,J.M., editors. Lippincott Williams & Wilkins, Philadelphia. 399-410.
- 38.) Diabetes Prevention Trial-Type 1 Diabetes Study Group. (2002) Effects of insulin in relatives of patients with type 1 diabetes mellitus. *N Engl J Med*. 346:1685-1691.
- 39.) Pugliese A, Zeller M, Fernandez A, Zalcberg L, Bartlett R, Ricordi C, Pietropaolo M, Eisenbarth G, Bennett S, Patel D. (1997). The insulin gene is transcribed in the human thymus and transcription levels correlates with allelic variation at the *INS* VNTR-IDDM2 susceptibility locus for type 1 diabetes. *Nature Genetics*. 15:293-297.
- 40.) Bennett S, Wilson A, Cucca F, Nerup J, Pociot F, McKinney P, Barnett A, Bain S, Todd J. (1996). *Journal of autoimmunity*. 9: 415-421.
- 41.) Mathews C, Pietropaolo S, Pietropaolo M. (2003). Reduced thymic expression of islet antigen contributes to loss of self tolerance. *Annals of the New York Academy of Sciences*. 1005: 412-417.
- 42.) Gerling I, Serreze D, Christianson S, Leiter E. (1992) Intrathymic islet cell transplantation reduces b-Cell autoimmunity and prevents diabetes in NOD/Lt Mice. *Diabetes*. 41:1672-1676.
- 43.) Posselt A, Barker C, Friedman A, Naji A. (1992) Prevention of autoimmune diabetes in the BB rat by intrathymic islet transplantation at birth. *Science*. 256:1321-1324.
- 44.) Pietropaolo M, Giannoukaki N, Trucco M. (2002) Cellular environment and freedom of gene expression. *Nat Immunol*. 3:335.
- 45.) Pugliese A, Brown D, Garza D, Murchison D, Zeller M, Redondo M, Diez J, Eisenbarth G, Patel D, Ricordi C. (2001) Self-antigen-presenting cells expressing diabetes-associated autoantigens exist in both thymus and peripheral lymphoid organs. *J Clin Invest*. 107:555-564.
- 46.) Chan S, Benoist C, Mathis D. (1994) The selective model of positive selection. *Semin Immunol*. 6:241-8.
- 47.) Sospedra M, Ferrer-Francesch X, Pujol-Borrell R, Juan M, Foz-Sala M. (1998) Transcription of a broad range of self-antigen in human thymus suggests a role for central mechanisms in tolerance toward peripheral antigens. *J Immunol*. 161:5918-5929
- 48.) Smith K, Olson, D, Hirose R, Hanahan D. (1997) Pancreatic gene expression in rare cells of thymic medulla: evidence for functional contribution to T cell tolerance. *International Immunol*. 9:1355-1365.

- 49.) Nepom G, Kwok W. (1998) Molecular basis for HLA-DQ association in IDDM. *Diabetes*. 47:1177-1184.
- 50.) Clare-Salzler M, Brooks J, Chai A, Van Herle K, Anderson C. (1992) Prevention of diabetes in nonobese diabetic mice by dendritic cell transfer. *J Clin Invest* 741-748.
- 51.) Feili-Hariri M, Dong X, Alber S, Watkins S, Salter R, and Morel P. (1999) Immunotherapy of NOD mice with bone marrow-derived dendritic cells. *Diabetes*. 48:2300-2308.
- 52.) Papaccio G, Nicoletti F, Pisanti F, Bendtzen K, Galdieri M. (2000) Prevention of spontaneous autoimmune diabetes in NOD mice by transferring *in vitro* antigen-pulsed syngeneic dendritic cells. *Endocrinology*. 141:1500-1505.
- 53.) Pearson T, Markees T, Wicker L, Serreze D, Peterson L, Mordes J, Rossini A, Greiner D. (2003) NOD Congenic Mice Genetically Protected From Autoimmune Diabetes Remain Resistant to Transplantation Tolerance Induction. *Diabetes*. 52:321-326.
- 54.) Bach J. (2001) Immunotherapy of insulin-dependent diabetes mellitus. *Curr Opin Immunol*. 13:601-605.
- 55.) Kukreja A, Cost G, Marker J, Zhang C, Sun Z, Lin-Su K, Ten S, Sanz M, Exley M, Wilson B, Porcelli S, Maclaren N. (2002) Multiple immuno-regulatory defects in type 1 diabetes. *J Clin Invest*. 109:131-140.
- 56.) Verge C, Gianani R, Kawasaki E, Yu L, Pietropaolo M, Jackson R, Chase P, Eisenbarth G. (1996) Prediction of type 1 diabetes in first-degree relatives using a combination of insulin, GAD and ICA51bdc/IA-1 Autoantibodies. *Diabetes*. 45:926-933.
- 57.) Pietropaolo M, Becker D, LaPorte R, Dorman J, Riboni S, Rudert W, Mazumdar S, Trucco M. (2002) Progression to insulin-requiring diabetes in seronegative prediabetic subjects: the role of two HLA-DQ high risk haplotypes. *Diabetologia*. 45:66-76.
- 58.) Nussbaum R, McInnes R, Willard H. (2004) Thompson & Thompson Genetics in medicine the 6<sup>th</sup> edition.130-132. Saunders-An Imprint of Elsevier. Philadelphia PA.
- 59.) Gilbert S, (2003). Developmental Biology seventh edition..688-689. Sinauer associates Inc. Sunderland MA.
- 60.) Canard B, Sarfati F. (1994) DNA polymerase fluorescent substrates with reversible 3' tags. *Gene*. 148:1-6.
- 61.) Pietropaolo, M, Yu S, Libman I, Pietropaolo S, Riley K, LaPorte R, Drash A, Mazumdar S, Trucco M, Becker D. (2005) Cytoplasmic islet cell antibodies remain valuable in defining risk of progression to type 1 diabetes in subjects with other islet autoantibodies. *Pediatr.Diabetes*. 6:184-192.

- 62.) Martin S, Kardorf J, Schulte B, Lampeter E, Gries F, Melchers I, Wagner R, Bertrams J, Roep B, Pflutzner A. (1995) Autoantibodies to the islet antigen ICA69 occur in IDDM and in rheumatoid arthritis. *Diabetologia* 38:351-355.
- 63.) Miyazaki, I, Cheung, R, Gaedigk, R, Hui, M, Van der Meulen, J, Rajotte, R, Dosch, H. (1995) T cell activation and anergy to islet cell antigen in type I diabetes. *J Immunol.* 154:1461-1469.
- 64.) Roep B, Duinkerken G, Schreuder G, Kolb H, De Vries R, Martin S. (1996) HLA-associated inverse correlation between T cell and antibody responsiveness to islet autoantigen in recent-onset insulin dependent diabetes mellitus. *Eur J Immunol.* 26:1285-1289.
- 65.) Dosch H, Cheung R, Karges W, Pietropaolo M, Becker D. (1999) Persistent T cell anergy in human type 1 diabetes. *J Immunol.* 163:6933-6940.
- 66.) Karges W, Hammond-McKibben D, Gaedigk R, Shibuya N, Cheung R, Dosch H. (1997) Loss of self-tolerance to ICA69 in nonobese diabetic mice. *Diabetes.* 46:1548-1556.
- 67.) Friday R, Jaquin-Gerstl A, Mathie T, Rudert W, Trucco M, Pietropaolo M. (1998) Potential role for thymic expression of the IDDM-related autoantigen ICA69 in central immune tolerance. *Am J Hum Genet* 63 (Suppl.):A181 (Abstr.).
- 68.) Song A, Winer S, Tsui H, Sampson A, Pasceri P, Ellis J, Elliot J, Dosch H. (2003) Deviation of islet autoreactivity to cryptic epitopes protects NOD mice from diabetes. *Eur J Immunol.* 33:546-555.
- 69.) Miyazaki I, Gaedigk R, Hui M, Cheung R, Morkowski J, Rajotte R, Dosch H. (1994) Cloning of human and rat p69, a candidate autoimmune target in type I diabetes. *Biochem Biophys Acta.* 1227:101-104.
- 70.) Hanahan D. (1998) Peripheral-antigen-expressing cells in thymic medulla: factors in self-tolerance and autoimmunity. *Curr Opin Immunol.* 10:656-662.
- 71.) Sospedra M, Ferrer-Francesch X, Pujol-Borrell R, Juan M, Foz-Sala M. (1998) Transcription of a broad range of self-antigen in human thymus suggests a role for central mechanisms in tolerance toward peripheral antigens. *J Immunol.* 161:5918-5929.
- 72.) Pugliese A, Brown D, Garza D, Murchison D, Zeller M, Redondo M, Diez J, Eisenbarth G, Patel D, Ricordi C. (2001) Self-antigen-presenting cells expressing diabetes-associated autoantigens exist in both thymus and peripheral lymphoid organs. *J Clin Invest.* 107:555-564.
- 73.) Jolicœur C, Hanahan D, Smith K. (1994) T-cell tolerance toward a transgenic beta-cell antigen and transcription of endogenous pancreatic genes in thymus. *Proc Natl Acad Sci USA.* 91:6707-67011.

- 74.) Pietropaolo M, Castano L, Babu S, Buelow R, Kuo Y, Martin S, Martin A, Powers A, Prochazka M, Naggert J. *et al.* (1993) Islet cell autoantigen 69 kDa (ICA69): Molecular cloning and characterization of a novel diabetes associated autoantigen. *J Clin Invest.* 92:359-371.
- 75.) Smith, K, Olson, D, Hirose, R, Hanahan, D. (1997) Pancreatic gene expression in rare cells of thymic medulla: evidence for functional contribution to T cell tolerance. *International Immunol.* 9:1355-1365.
- 76.) Heath V, Moore N, Parnell S, Mason D. (1998) Intrathymic expression of genes involved in organ specific autoimmune diseases. *J Autoimmun.* 11:309-318.
- 77.) Bach J, Chatenoud L. (2001) Tolerance to islet autoantigens in type 1 diabetes. *Annu Rev Immunol.* 19:131-161.
- 78.) Pugliese A, Diez J. (2002) Lymphoid organs contain diverse cells expressing self-molecules. *Nat Immunol.* 3:335-336.
- 79.) Spitzenberger F, Pietropaolo S, Verkade P, Habermann B, Lacas-Gervais S, Mziaut H, Pietropaolo M, Solimena M. (2003) Islet Cell Autoantigen of 69 kDa is an arfaptin-related protein associated with the Golgi complex of insulinoma INS-1 cells. *J Biol Chem.* 278:26166-26173.
- 80.) Vafiadis P, Bennett S, Todd J, Nadeau J, Grabs R, Goodyear C, Wickramasinghe S, Colle E, Polychronakos C. (1997) Insulin expression in the thymus is modulated by *INS* VNTR alleles at the *IDDM2* locus. *Nat. Genet.* 15:289-292.
- 81.) Werdelin O, Cordes U, Jenson T. (1998) Aberrant expression of tissue-specific proteins in the thymus: a hypothesis for the development of central tolerance. *Scand J Immunol.* 47:95-100.
- 82.) Throsby M, Homo-Delarche F, Chevenne D, Goya R, Dardenne M, Pleau J. (1998) Pancreatic hormone expression in the murine thymus: localization in dendritic cells and macrophages. *Endocrinology* 139:2399-2406.
- 83.) Todd J. (1995). Genetic analysis of type 1 diabetes using whole genome approaches. *Proc. Natl. Acad. Sci.* 92: 8560-8565.
- 84.) Barnette T, Gourrand P, Cambon-Thomsen A. (2005) Strategies in analysis of the genetic component of multifactorial diseases; biostatistical aspects. *Transplant Immunology.* 14 255-266.
- 85.) Friday R, Pietropaolo S, Profozich J, Trucco M, Pietropaolo M. (2003) Alternative core promoters regulate tissue-specific transcription from autoimmune diabetes-related *ICA1* (ICA69) gene locus. *The Journal of Biological Chemistry.* 278:853-863.

- 86.) Gaedigk R, Karges W, Hui M, Scherer S, Dosch H. (1996) Genomic organization and transcript analysis of ICAp69 a target antigen in diabetic autoimmunity. 38:382-391. *Genomics*.382-391.
- 87.)Pietropaolo M, Castano L, Babu S, Buelow R, Kuo Y, Martin S, Martin A, Powers A, Prochazka M, Naggert J, Leiter E, Elsenbarth G. (1993) Islet cell autoantigen 69kD (ICA 69) - Molecular cloning and characterization of a novel diabetes-associated autoantigen. *J. Clin. Invest.* 92:359-371.
- 88.) Bart R. (1996) T-cell responses to autoantigens in IDDM: The search for the holy grail. *Diabetes*. 45:1147-1156.
- 89.) Dow E, Gelding S, Skinner E, Hewit J, Gray I, Mather H, Williamson R, Johnston D. (1994) Genetic analysis of glucokinase and the chromosome 20 diabetes susceptibility locus in families with type 2 diabetes. *Diabet Med*. 11: 856-861.
- 90.) Cook J, Hattersley A, Christopher P, Bown E, Barrow B, Patel P, Shaw J, Cookson W, Permutt M, Turner R. (1992) *Diabetes*. 41:1496-1500.
- 91.) Lotif K, Sund G, Lowe R, Graham J, Landin-Olsson M, Kockum I, Deeb S, LernmarkA, the diabetes incidence study in Sweden. (1997) The beta cell glucokinase promoter variant is an unlikely risk factor for diabetes mellitus. *Diabetologia*. 40: 959-962.
- 92.) Pilon M, Peng X, Spence A, Plasterk R, Dosch H. (2000) The diabetes autoantigen ICA 69 and its *canorhabditis elegans* homologue, *ric-19*, are conserved regulators of neuroendocrine secretion. *Molecular biology of the cell*. 11:3277-3288.
- 93.) Winer S, Gunaratnam L, Astsatourov I, Cheng R, Kubiak V, Karges W, Hammond-McKibben D, Gaedigk R, Graziano D, Trucco M, Becker D, Dosch H . (2000) Peptide dose, MHC affinity, and target self-antigen expression are critical for effective immunotherapy of nonobese diabetic mouse prediabetes. *The Journal of Immunology*. 165:4086-4094.
- 94.) Karges W, Hammond-McKibben D, Gaedigk R, Shibuya N, Cheng R, Dosch H. (1997) Loss of self-tolerance to ICA69 in nonobese diabetic mice. *Diabetes*. 46:1548-1556.
- 95.) Kishimoto H, Sprent J. A defect in central tolerance in NOD mice. (2001) *Nat Immunol*. 2:1025-1031.
- 96) Corcoran D, Feingold E, Cominick J, Wright M, Harnaha J, Trucco M, Giannoukakis N, Benos P. (2005) Footer: a quantitative comparative genomics method for efficient recognition of cis-regulatory elements. *Genome Research*. 15: 840-847.
- 97.) Sanger F, Nicklen S, Coulson A. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad Sci*. 74: 5463-5467.



- 98.) Hyman E. (1988) A new method of sequencing DNA. *Anal Biochem.* 174:423-436.
- 99.) Ronaghi M, Uhlen M, Nyren P. (1998) A sequencing method based on real-time pyrophosphate. *Science.* 281: 363-365.
- 100.) Rosenthal A, International Patent Application Publication 761107 (1996).
- 101.) Metzker M, Raghavchari R, Richards S, Jacutin S, Civitello A, Burgess K, Gibbs R. (1994) Termination of DNA synthesis by novel 3'-modified-deoxyribonucleosides 5'-triphosphates. *Nucleic Acid Res.* 22:4259-4267.
- 102.) Jones D. (1997) An iterative and regenerative method of DNA sequencing. *Biotechniques.* 22:938-946.
- 103.) Ronaghi M. (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Research.* 11:3-11.
- 104.) Cheesman P. (1994) Method for sequencing polynucleotides. US patent no. 5302509.
- 105.) Tsien R, Ross P, Fahnestock M, Johnston A. (1991) Method for DNA sequencing. US patent no PCT WO 91/06678.
- 106.) Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P. (1996) Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* 242:84-89.
- 107.) Alexander A, Pecoraro C, Styche A, Rudert W, Benos P, Ringquist S, Trucco M. (2005) SOP3: a web-based tool for selection of oligonucleotide primers for single nucleotide polymorphism analysis by pyrosequencing. *BioTechniques* 28: 87-94.
- 108.) Ahmadian A, Lundeberg J, Nyren P, Uhlen M, Ronaghi M. (2000) Analysis of the p53 tumor suppressor gene by pyrosequencing. *BioTechniques.* 28:140-144.
- 109.) National Diabetes Data Group. (1979) Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes* 28:1039-1057.
- 110.) Biotage (2005) <http://www.pyrosequencing.com/>
- 111.) Bassuny W, Ihara K, Sasaki Y, Kuromaru R, Kohno H, Matsuura N, Hara T. (2003) A functional polymorphism in the promoter/enhancer region of the FOXP3/Scurfin gene associated with type 1 diabetes. *Immunogenetics.* 55:149-156.
- 112.) HapMap- <http://www.hapmap.org/>
- 113.) Lie B, Ronningen K, Akselsen H, Thorby E, Undlien D. (2000) Application and interpretation of transmission/ disequilibrium tests: transmission of HLA-DQ haplotypes to unaffected siblings in 526 families with type 1 diabetes. *Am. J. Hum. Genet.* 66: 740-743.

- 114.) Spielman R, McGinnis R, Ewens W. (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* 52: 506-516.
- 115.) Kaplan N, Martin E, Weir B. (1997) Power studies for the transmission/ disequilibrium tests with multiple alleles. *Am. J. Hum. Genet.* 60:691-702.
- 116.) Mimbacas A, Perez-Bravo F, Santos J, Pisciotto C, Grignola R, Javiel G, Jorge A, Cardosos H. (2004). The association between HLA DQ genetic polymorphism and type 1 diabetes in a case-parent study conducted in admixed population. *Europ. J. O Epidem.* 19: 931-934.
- 117.) Nobel J, Valdes A, Bugawan T, Apple R, Thomson G, Erlich H. (2002) The HLA class I A locus affects susceptibility to type 1 diabetes. *Human Immun.* 62. 657-664.
- 118.) Guja C, Guja L, Nutland S, Rance H, Sebastien M, Todd J, Ionescu-Tirgoviste C. (2004) Type 1 diabetes genetic susceptibility encoded by HLA DQB1 genes in Romania. *J. Cell. Mol. Med.* 2: 249-256.
- 119.) Ranade K, Change M, Ting C, Pei D, Hsiao C, Oliver M, Pesich R, Hebert J, Chen Y, Dzau V, Curb D, Olshen R, Risch N, Cox D, Bostein. (2001) High-throughput with single nucleotide polymorphisms. *Genome Research.* 11:1262-1268.